

A PSYCHOMETRIC EVALUATION OF THE
CORRECTIONAL ADJUSTMENT CHECKLIST

BY

BRAINARD WILLEM HINES

A DISSERTATION PRESENTED TO THE GRADUATE
COUNCIL OF THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DOCTOR OF PHILOSOPHY DEGREE

UNIVERSITY OF FLORIDA

1980

ACKNOWLEDGMENTS

My sincerest thanks go to the members of my doctoral committee for their patience and understanding. Particularly, I would like to thank the chairman of my committee, Dr. William Ware who has been a good friend and constructive influence throughout my academic career at the University of Florida. I also owe a special debt of gratitude to Dr. Linda Crocker and Dr. Richard Swanson for their support and encouragement.

In addition, I would like to express my gratitude to the University of North Carolina at Chapel Hill for allowing Dr. Ware to continue as my chairman for the past months.

Finally, I would like to express my appreciation to my wife Magdalena Llabre, whose love and encouragement made this dissertation possible.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGMENTS	iii
LIST OF TABLES	v
ABSTRACT	vii
 Chapter	
I. INTRODUCTION	1
Psychometric Properties Investigated	4
Definitions of Reliability	4
Definitions of Validity	5
Construct Validation	7
Statement of the Problem	8
Significance of the Study	10
II. REVIEW OF THE LITERATURE	12
Classification	13
Classification of Criminals	15
Empirically Derived Typologies	19
Current Reviews of Criminal Typologies	21
Psychometric Concepts	30
Reliability Estimation in This Study	37
Validity	41
Types of Validity	44
Construct Validity Estimates	46
Chapter Summary	51
III. METHOD	53
The Sample	53
Selection of the Sample	55
Instrumentation	59
Data Collection	66
Data Analysis	69
Reliability of the CACL	69
Predictive Validity of the CACL	71

Construct Validation of the CACL	73
Postdiction of Crime Type	74
Summary	75
IV. RESULTS	77
Inter-Rater Reliability of the CACL	78
Construct Validation of the CACL	79
Criterion-Related Validity of the CACL	85
Relationship of the CACL to Crime	91
Summary	93
V. DISCUSSION	94
The Inter-Rater Reliability of the CACL	96
Construct Validation of the CACL	96
Criterion Validity of the CACL	101
Suicide Attempts	102
Threats of Assault	104
Assaults	105
Infractions of Rules	105
Relation of the CACL to Crime Type	106
Summary of Psychometric Evaluation	107
Recommendations	110

Appendix

A. CORRECTIONAL ADJUSTMENT CHECKLIST	113
B. SUMMARY TABLES FOR INTER-RATER RELIABILITY STUDIES	116
REFERENCES	123
BIOGRAPHICAL SKETCH	132

LIST OF TABLES

TABLE	PAGE
1. NUMBER OF RESIDENTS BY UNIT ADMITTED TO NFETC FROM ITS INCEPTION UNTIL JULY 1, 1978	58
2. DESCRIPTIVE STATISTICS FOR CONCURRENT VALIDITY STUDY	80
3. INTERCORRELATION MATRIX FOR CONCURRENT VALIDATION STUDY	81
4. RESULTS OF CANONICAL CORRELATION ANALYSIS OF THE CACL AND MMPI	83
5. CANONICAL WEIGHTS OF MMPI AND CACL SUB- TESTS FOR CANONICAL VARIATES 1 AND 2	84
6. PRODUCT MOMENT CORRELATION BETWEEN SUB- TESTS OF THE CACL AND MMPI AND CANON- ICAL VARIATES	86
7. DESCRIPTIVE STATISTICS FOR PREDICTIVE VALIDITY STUDY	87
8. INTERCORRELATION MATRIX FOR PREDICTIVE VALIDITY STUDY	88
9. RESULTS OF MULTIPLE REGRESSION OF FREQUENCY OF SUICIDE ATTEMPTS ON CACL SUBSCALES	89
10. RESULTS OF MULTIPLE REGRESSION OF FREQUENCY OF ASSAULTS ON CACL SUB- SCALES	89
11. RESULTS OF MULTIPLE REGRESSION OF FREQUENCY OF THREATS OF ASSAULT ON CACL SUBSCALES	90
12. RESULTS OF MULTIPLE REGRESSION OF FREQUENCY OF INFRACTIONS ON CACL SUBSCALES	90

13.	RESULTS FOR DISCRIMINANT FUNCTION ANALYSIS OF CRIME TYPE	92
-----	-----------------------------------------------------------------------	----

APPENDIX

14.	DESCRIPTIVE STATISTICS FOR INTER- RATER RELIABILITY STUDY: "INTAKE" CONDITION	116
15.	DESCRIPTIVE STATISTICS FOR INTER- RATER RELIABILITY STUDY: "CONTROLLED" CONDITION	116
16.	ANALYSIS OF VARIANCE TABLE FOR CACL PA "CONTROLLED" CONDITION INTER- RATER RELIABILITY STUDY	117
17.	ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL ID "CONTROLLED" CONDITION INTER-RATER RELIABILITY STUDY	117
18.	ANALYSIS OF VARIANCE TABLE FOR CACL NA "CONTROLLED" CONDITION INTER- RATER RELIABILITY STUDY	118
19.	ANALYSIS OF VARIANCE TABLE FOR CACL MA "CONTROLLED" CONDITION INTER- RATER RELIABILITY STUDY	118
20.	ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL PA "INTAKE" CONDITION INTER-RATER RELIABILITY STUDY	119
21.	ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL ID "INTAKE" CONDITION INTER-RATER RELIABILITY STUDY	119
22.	ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL NA "INTAKE" CONDITION INTER-RATER RELIABILITY STUDY	120
23.	ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL MA "INTAKE" CONDITION INTER-RATER RELIABILITY STUDY	120
24.	INTER-RATER RELIABILITY COEFFICIENTS FOR INTAKE AND CONTROLLED CONDITIONS, INCLUDING SYSTEMATIC RATER BIAS IN THE ERROR TERM	121

Abstract of Dissertation Presented to the
Graduate Council of the University of Florida
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

A PSYCHOMETRIC EVALUATION OF THE CORRECTIONAL
ADJUSTMENT CHECKLIST

BY

Brainard Willem Hines

June 1980

Chairman: Professor William B. Ware
Major Department: Foundations of Education

A variety of classification systems have been developed for use in the social sciences. These systems have become increasingly complex with the advent of modern statistical techniques. The use of classification systems in the field of criminology began with the effort to discriminate between criminals and "normals" on the basis of physical features. Although more recent classification systems in the field of criminology attempt to define types of criminals or criminal behavior, few have been adequately evaluated in terms of the psychometric properties of reliability and validity.

The Correctional Adjustment Checklist (CACL) is a factor-analytically derived classification instrument which is designed to describe the behavior of incarcerated males

along four dimensions. These dimensions have been labelled Psychopathic-Aggressive (PA), Neurotic-Anxious (NA), Immature-Dependent (ID), and Manipulative (Ma). Ratings along these dimensions are intended to have differential implications for the management and treatment of individuals in close confinement. Although the instrument has been used in a variety of settings, little information is available on its inter-rater reliability or validity.

This study attempted to evaluate the CACL by assessing the degree of congruence among raters in a naturalistic setting and under conditions designed to provide maximal reliability estimates. Data gathered under both conditions provided reliability estimates for the average of three raters which ranged upwards of .60, with the exception of the Ma subscale, which showed a lower inter-rater reliability estimate in the "controlled" condition.

Assessment of the validity of the CACL with this example involved estimating the relationship between it and other variables of interest. In this study, these variables were as follows: scores on the MMPI administered concurrently with the CACL; the frequency of several types of disruptive behavior during the first sixty days after the CACL was administered; and the degree of violence involved in the crime with which the subjects had been most recently charged.

These estimates of the CACL's relationship with other variables are statistically significant in several instances. First, a canonical variate analysis derived two sets of variables from the CACL and MMPI. Although both canonical correlation coefficients are significant at the .05 level, a redundancy analysis indicates that the relationship between the two instruments is very modest. Also, scores on the CACL showed a statistically significant relationship to suicide attempts and threats of violence which occurred within the first sixty days.

When the subscales of the CACL were used as the predictors in a multiple regression analysis, the NA subscale showed the highest degree of association with suicide attempts, followed by the PA subscale. Additionally, the PA subscale is the only subtest which accounts for a statistically significant amount of variance in verbal threats of physical violence. Other disruptive behaviors (actual assaults and other infractions) were not significantly related to scores on any of the CACL's subscales.

A discriminant function analysis did not show any significant relationship between subscale scores on the CACL and the presence of physical violence in the subjects' most recent crime. This may have been due to the imprecise match between the charge (e.g., armed robbery) and the actual degree of violence in the crime.

In summary, the CACL provides subscale scores which are reliable across raters, and which predict several behaviors of interest within a maximum security mental hospital setting. It shows a modest degree of redundancy with the MMPI, indicating that it may well be measuring factors not being tapped by that instrument. Although the CACL was developed for the classification of a general prison population, it appears to have utility when used with individuals who are emotionally disturbed.

.

CHAPTER I

INTRODUCTION

The process of classification of objects or events is essential to the development of any science. Although the origin of taxonomy or classification goes back to the ancient Greeks, the advent of modern statistical techniques and the use of high-speed computers have allowed for the use of more sophisticated classification methods than have been previously possible. In many areas, such as entomology, the development of more sophisticated taxonomic systems has contributed to the general advancement of the field in question.

The use of classification in criminology can be traced to a number of physiologists such as Lambroso who, in the nineteenth century, attempted to define a "criminal type" on the basis of physical features. Such typologies were intended to discriminate between criminals and "normals," not to classify types of individuals who had been convicted of crimes nor relate those types to other measures of any sort.

With the modern emphasis on treatment rather than custody of criminals has come a concern for possible subtypes of offenders. Additionally, the more

humanitarian philosophy of our own era has encouraged the scientific study of criminal behavior and the types of individuals who become criminals. Such efforts to create meaningful offender typologies (Gibbons, 1975) have come about because of the failure of unitary treatment approaches and because of the observed variance in types of crimes, demographic and personal or behavioral characteristics of criminals.

Although a variety of offender typologies have been proposed, none has been widely accepted. Many typologies either are based on traditional psychological personality types or are concerned with classifying offenders based on the type of crime which they have committed. Other classification systems have been impressionistic and have included a variety of types which have not been found by other investigators (Gibbons, 1975). Generally, no single typology of criminals or criminal behavior has been found to be of use in a variety of settings or with various age groups of individuals. Also, no single typology has been constructed which is of use in delineating both the etiology and diagnostic category of criminal behavior.

Most criminal typologies are based on the results of a single instrument or a series of descriptions of the crime or its etiology. Few classification systems used in criminology are based on empirically derived methods, but rather are derived from theoretical formulations. Quay

(1971) has made one of the few attempts to empirically construct a classification system for criminals.

The Quay Correctional Adjustment Checklist (CACL) is an instrument derived using factor analysis for the purpose of describing the behavior of incarcerated individuals on four dimensions (Quay, 1971). These dimensions are labelled Psychopathic-Aggressive (PA), Neurotic-Anxious (NA), Immature-Dependent (ID), and Manipulative (Ma). It is intended not only to describe an individual's patterns of behavior within an institution, but also to provide information useful for differential treatment based on those patterns.

Although the CACL has been used in a variety of settings, its psychometric properties have never been thoroughly investigated. A review by Warren (1969) reported that the CACL appeared to have "adequate" reliability but gave no source for that statement. Another article by Quay (1971) has called for further study of the Checklist, but gave no validity estimates for the instrument.

The purpose of this study is to investigate the inter-rater reliability and the validity of this instrument, based on the behavior of a sample of individuals who have been confined in a maximum security mental hospital in Gainesville, Florida. All of those individuals have either been convicted of a felony, have been found incompetent to stand trial for a felony, or are not guilty

by reason of insanity. Data on the CACL have never been gathered on a psychiatric population (Quay, personal communication, 1978). If adequate reliability and validity estimates are obtained for the sample, the CACL should be used in other such settings.

Psychometric Properties Investigated

Definitions of Reliability

As Kerlinger and Pedhazur (1973) pointed out, there are a variety of definitions of reliability. In general, they defined reliability as the consistency and accuracy of an instrument which are related to the absence of random or error variance in that instrument. Specifically, he wrote that ". . . reliability can be defined as the relative absence of errors of measurement in a measuring instrument" (p. 443).

The reliability of any measure can be thought of as existing in any of several dimensions. These dimensions may involve consistency (freedom from measurement error) across time, across items in a single measure, across other forms of the test or across raters or scorers on the same form of the test. The types of reliability which correspond to the degree of consistency in each of these dimensions are known as test-retest (time), internal consistency (items), parallel forms (forms) and inter-rater reliability.

The central focus of this study is the consistency of the CACL scores across raters who have observed the

individual under similar circumstances and who have received similar training in the use of the instrument. This type of reliability (inter-rater) is of paramount importance to the CACL, since it is intended to measure the presence of observable behaviors. If equally trained observers cannot agree on whether a particular behavior is present, then usefulness of this instrument for any practical purpose is highly questionable.

Definitions of Validity

The Standards for Educational and Psychological Tests and Manuals (1974), published by the American Psychological Association, stated:

Validity information indicates the degree to which the test is capable of achieving certain aims. Tests are used for several types of judgment, and for each type of judgment a different type of investigation is required to establish validity.
(p. 13)

That is, validation is defined as a process or activity performed on the data arising from a test. The manner in which the data are treated is intended to parallel an aspect of the intended use of the measure, or of its interpretability. The Standards publication goes on to list three aims of testing which correspond to three types of validation procedures.

1. The test user wishes to determine how an individual performs at present in a universe of situations that the test situation is claimed to represent.

2. The test user wishes to forecast an individual's future standing or to estimate an individual's present standing of some variable of particular significance that is different from the test.
3. The test user wishes to infer the degree to which the individual possesses some hypothetical trait or quality (construct) preserved to be reflected in the test performance.
(p. 13)

The American Psychological Association and the American Educational Research Association have defined three basic types of validity: content, construct, and criterion-related (APA, 1974). Among these three types, criterion-related and construct validity are most appropriate in assessing the potential usefulness of the CACL.

Criterion-related validity encompasses both predictive and concurrent validity which reflect the correlation between scores on a test and performance on a criterion variable. In concurrent validation, measures on both test and criterion are obtained at approximately the same point in time; predictive validation occurs when the criterion measure is taken after the test in question. For this study, predictive criterion-related validity will be investigated by relating CACL scores at the time of intake to the frequencies of several types of disruptive behavior, recorded during the first sixty days of confinement at the hospital.

Construct Validation

Construct validation is a complex procedure attempting to ascertain the degree to which a measure empirically relates to a number of other variables which logically and deductively derive from the construct which the instrument purports to measure. To do this, a construct validation study often involves an attempt to demonstrate that the trait is related to other variables which are logically inherent from the construct, and that variables which do not logically derive from the construct do not empirically relate to it.

In the same way, this study will assess the degree to which the CACL's subscales relate to other variables which logically derive from the constructs they purport to measure. That is, we would expect that individuals who score highly on the Psychopathic-Aggressive subscale would be more violent and disruptive than those who score highly on the Immature-Dependent subscale. In addition, they should more often take a leadership role and rely less on staff for advice than other types of individuals. Also, we would expect such individuals to be more frequently threatening to others, and to score more highly on those subscales of another instrument which measure impulsivity and hostility. If such relationships were evident, this would help in defining the nature of the basic traits being assessed by the CACL.

The analyses conducted in this study provide estimates of the relationship between the CACL and several behaviors which are of interest in the institution. They also provide an estimate of the nature and degree of relationship between the CACL and the Minnesota Multiphasic Personality Inventory, a self-report diagnostic instrument which has been used in other criminal classification systems.

Statement of the Problem

This study will be addressed to determining the psychometric properties of the Quay Correctional Adjustment Checklist based on the performance of a sample of individuals in a maximum security mental hospital. It will not address the decision rules used to classify individuals, but will deal only with the psychometric properties of the instrument. Specifically, it will attempt to answer the following questions:

1. What is the degree of agreement among raters with similar training who rate individuals on each of the four subscales of the CACL? (Inter-rater reliability)
2. What is the degree of association between the subscales on the CACL and the subscales of the MMPI, when both instruments are administered concurrently? (Construct validity)
3. Is there a relationship between the various subscales of the CACL and the type of crime which caused the individuals' incarceration? (Construct validity)

4. What is the relationship between scores on the CACL and an index of disruptiveness within the institution? (Criterion-related and Construct validity)

The questions above are concerned respectively with inter-rater reliability and validity, both of which are important to the use of the CACL in institutional settings. It is important to note that the study is not designed to explore the use of particular decision rules for classification. Rather, it is concerned with the consistency and "interpretability" of scores on the CACL. Also, the reliability estimates which are given are for the average of three raters, where each rater rates all individuals. These estimates are considerably higher than those which would be obtained for a single rater.

At a time when there is a clear need for adequate classification techniques in the area of corrections (Warren, 1969), the CACL presents some unique advantages and disadvantages. Although it has been used in a variety of settings, often for the purpose of classification for treatment, its psychometric properties remain for the most part unknown. If its reliability and validity are low, its use should be discontinued. If they are acceptable, the instrument's utility could be explored in other settings. These possibilities are further discussed in the next section.

Significance of the Study

As Gibbons (1975) has pointed out, there is increasing dissatisfaction with the process of classification in criminology and criminal justice. Many of the current problems with the medical model, often used in criminal classification, may well be due to the ineffectiveness of current diagnostic instruments and the consequent misclassification of many individuals. Despite the apparent failure of treatment strategies which presume a single type of offender, no type of classification system other than the CACL has evolved which is based on actual patterns of behavior in an institution.

Quay (1971) noted that: "Additional research with respect to reliability and construct validity (of the CACL) is in order" (p. 11). Although this need has been recognized since the initial development of the instrument, such studies have not been forthcoming. Despite the fact that there has never been adequate assessment of the instrument's psychometric properties, the CACL has been used in a variety of institutions, including the Robert Kennedy Federal Youth Center in Morgantown, West Virginia, the North Florida Evaluation and Treatment Center in Gainesville, Florida, and the Federal Correctional Institution in Miami, Florida.

Studies such as this one are important for several reasons. First, if the instrument misclassifies offenders,

it may be hindering their effective treatment. Such misclassification is unfair to individual offenders and to the society which supports such treatment efforts. Second, the continued use of an instrument with unknown psychometric properties may well contribute to the increasing disenchantment with differential treatment strategies for offenders. Third, although the instrument may have an important function in the derivation of new theories in criminology, such functions will be of little use until validity is established. Additionally, the instrument may have potential use in the assessment of treatment effectiveness for individuals as well as groups of offenders. If it provides a reliable and valid measure of institutional behavior, it could allow for improved monitoring of those behavioral changes which occur during incarceration.

In this chapter an overview of classification as a process has been presented,⁷ and the application of this process to the field of criminology has been summarized. The psychometric properties of the CACL, which are investigated in this study, are summarized along with the implications of the study. The following chapter includes a review of the literature on classification as a logical process, its use in criminology, and on the psychometric properties of reliability and validity.

CHAPTER II

REVIEW OF THE LITERATURE

This study is intended to provide reliability and validity estimates for the CACL, based on the behavior of a sample of individuals incarcerated in a maximum security mental hospital. Since the CACL is an instrument intended to classify criminals, the literature review first includes a summary of articles on classification as a field of study in its own right. This is followed by a more extensive review of the development of criminal typologies. The trend towards empirically developed rather than theoretically oriented systems is discussed. Next, the development of the CACL is outlined, and the instrument is compared with other classification systems which are based on self-report instruments rather than behavioral observation. The need for psychometric evaluation of the CACL is pointed out, and the importance of inter-rater reliability and further validation studies is emphasized. The final section of the literature review provides a brief review of the theoretical definitions of reliability and validity as they pertain to the CACL. The need for consistent rating of individuals is stressed, along with the need for meaningfulness and utility in the classifications which are derived from the instrument's

use. Thus, the inter-rater reliability and construct validity of the CACL are the areas of primary interest which are explored in this study. Although the criterion-related validity of the CACL is also investigated, the relationship of the criterion variables to the constructs measured by instrument is also explored.

Classification

Classification may be defined as the arrangement of objects or events into sets on the basis of their common characteristics. This process has been part of the natural sciences for centuries, but has only recently become a field of study in its own right. As such, the term taxonomy has been used to mean the theoretical study of classification as it occurs in a variety of specific disciplines.

One of the most comprehensive reviews of taxonomy appeared in 1974 in which Sokal reviewed the purposes, development, and structure of any classificatory activity. This review covered several general areas which are applicable to the use of classification in criminology, particularly the criteria for a desirable taxonomic system and the major purposes of classification.

Sokal (1974) said that,

The paramount purpose of a classification is to describe the structure and relationship of the constituent objects to each other and to simplify these relationships in such a way that general statements can be made about classes of events.
(p. 1116)

Implicit in this definition are several purposes of taxonomy. First, classification may be used to reveal the "true" relationships between objects or events by ordering them on the basis of common characteristics. Second, classification can be used to achieve economy of memory. By grouping single cases it provides the capacity to summarize information and to avoid repetition. Third, classification provides for ease of manipulation and facilitates information retrieval. It may be used to simplify problems in routing or delivery, to define political districts or to allow for cataloging printed materials. Finally, Sokal noted that classification systems have the primary scientific purpose of generating hypotheses, in that they should "stimulate interest as a means of furthering investigation" (p. 1117).

Sokal made several important points about the purpose and types of classification systems. He noted that classification systems may serve the purpose of economy of memory, reveal "natural" relationships between elements in each taxon, provide for ease of manipulation, and generate interest in new scientific problems. Classification systems may vary in the number of salient dimensions which they include and may be monothetic or polythetic in nature, depending on whether the elements in each taxon must share a common trait (in the former case) or whether an element may possess any combination of the traits (in the latter case).

In general, Sokal pointed out that the classification is emerging as a distinct discipline and that a "meta-typology" or classification of classification systems is possible. An ideal classification system should accommodate all elements of the set of objects or events to be classified, and should enable the typologist to match the dimensions and specificity of the system to its intended use. These criteria will be used later in this study to evaluate the CACL as a typological instrument in the field of criminology.

Classification of Criminals

An excellent overview of the classification of criminals is provided by Schafer (1968), who not only provided a historical narrative of the major typologists, but also defined several categories of criminal typologies (pp. 143-144). These include legal typologies of crime type, multiple cause typologies, typologies based on sociological or psychological theories, typologies which stress physiological factors, and those which describe the longitudinal development of criminal behavior.

Included in this section are an overview of current criminal typologies which fall into these categories and an explanation of the relationship between the CACL and these typologies. Generally, it is of interest that the CACL does not fall readily into any of Schafer's categories, since it deals with the behavior of felons while

incarcerated rather than the longitudinal development of criminal behavior.

The categories developed by Schafer emphasize either the hypothetical "causes" of criminal behavior, the type of crime(s) committed, or attempt to relate the two in a single description of a criminal "role career." Although some typologies are most often used for reporting frequencies of particular criminal acts in a given geographic area and are thus primarily empirical, most of the other typologies reflect on underlying theory of the causes of criminal behavior.

Legal typologies represent monothetic classification systems in which crimes rather than criminals are classified. The FBI Uniform Crime Reports for the United States represent such a system. Schafer noted that although such systems have historical and legal interest, "They are technical divisions for the use of the administration of justice and are not conceived of as explanations for behavior" (p. 146). Such a typological system will be used in this study to relate crime types to CACL classifications.

Multiple-cause typologies stress the interaction of biological, social and psychological causes of criminal behavior. Such systems were first developed in Germany in the nineteenth century by theorists who emphasized the affective and motivational components of criminal behavior as they exerted their influence across the criminal's life

span. This historical perspective was later used by Gibbons (1965, 1970) in his typology of criminal role careers. Gibbons, along with Clinkard and Quinney (1967) based a major criminological text on a multiple-cause typology. However, in a later article, Gibbons (1975) has emphasized the difficulties in the use of such a typology.

Sociological and psychological typologies both emphasize the hypothetical causes for criminal behavior. Sociological typologies attempt to delineate the external forces which contribute to criminal behavior, while psychological classification systems reflect the inner dynamics which may lead to such acts. Although these typologies reflect some of the most productive and extensive areas of offender classification, they also elicit some of the more vehement opposition (Schafer, 1968, p. 155).

Two of the most prominent sociological typologies are those of Tappan (1967) and Thrasher (1963). These two individuals have attempted to relate criminal behavior to its social causes, but frequently derived hypothetical multiple factor typologies with little empirical verification (Schafer, 1968). This is a general problem with etiological typologies, since they limit validation studies to ex post facto designs.

Psychological typologies are exemplified by the work of Alexander and Staub (1956) and Abrahamson (1960). Such

systems suffer from some of the same problems as those depending on more sociological explanations, since they are limited to the assessment of current psychological functioning. Current functioning in criminals may not reflect the dynamics in operation at the time the crime was committed.

Constitutional typologies have the most lengthy history in criminology, dating to Galen (circa, 150 A.D.). This group of typologies centers around the biopsychological causes of crime, especially the morphology of the offenders. The works of Lambroso (1911), Kretschmer (1925), and Sheldon (1949, 1954) are typical of this area.

Although these authors consistently have tried to relate body type to crime type, George Vald (1958) points out that "there is no present evidence at all of physical type, as such, having any consistent relation to legal and sociologically defined crime" (p. 129). Thus, constitutional typologies may⁷ allow for the classification of criminals, but the resultant categories have no empirical relationship to any manifest behavior. This problem is not unique to constitutional typologies, as will be shown later.

Normative typologies attempt to define the criminal's total personality in an effort to identify the "types" for which a particular sentence is appropriate. As such they incorporate a variety of legal, sociological, and psychological typologies. German authors have been primarily responsible for work in this area which has been little used in America.

Life-trend typologies are similar to multiple-facet typologies, but stress the dynamic structural coherence of the individual criminal's way of life. They are typically more complex than Gibbon's "role-careers" in that they attempt to follow the criminal behavior which is not part of a criminal life style.

Authors such as Reckless (1967) and Clinkard (1963) have developed systems of this type and have generally made a large impact in the field of criminology (Schafer, 1968, p. 6). This may well be due to the comprehensiveness of the system itself and the polythetic, multi-dimensional process which they use to classify offenders.

Empirically Derived Typologies

Given the problem inherent in the classification systems previously discussed, individuals such as Quay (1964), Gibbons (1975) and Megargee (1977) have recommended the use of empirically derived typologies. In such systems criminals are grouped on the basis of current behavior or demographic variables, without first theorizing about the causes for antisocial behavior. Accordingly, the development of such typologies differs from that involved in more "theory-oriented systems."

The effort to develop empirically keyed classificatory systems involves the administration of an instrument to a group of offenders and the development of a classificatory system based on the results of that instrument.

The important distinction to be made is that such typologies assess the current responses of the individual and are of limited scope and purpose. They are designed primarily for their immediate rather than long-term utility value, and may not have a predetermined underlying construct which they attempt to measure. Examples of such instruments are the classificatory subscales of the MMPI developed by Pantou (1965, 1966, 1968, 1970); Quay's work on the CACL (1971); and Megargee's recent work (1977), which also uses items on the MMPI.

These classificatory techniques have several common elements: first, they are not based on a single etiological or explanatory construct; second, they use a single instrument or subscale of an extent measure; and third, they describe current levels of functioning of the individual. These instruments are usually constructed by relating items to an external criterion (i.e., behavior in the institution) or to an internal criteria of factorial homogeneity, as is the case with the CACL and the Megargee MMPI system.

The CACL was developed by Quay as part of such an empirically derived classification system. Basing his classification system on the techniques developed by Hewitt and Jenkins (1946), Quay (1971) developed instruments assessing both current functioning (the CACL) and life history (CALH).

Although the development of the CACL will be discussed in greater detail in the next chapter, it should be mentioned here that the CACL is a behavioral checklist intended to assess patterns of current functioning while incarcerated. The instrument was normed on a prison population, and provides normalized T scores in each of four dimensions: Psychopathic-Aggressive (PA); Neurotic-Anxious (NA); Immature-Dependent (ID); and Manipulative (Ma). It is of interest that the CALH, a life-history checklist also groups criminals into these categories, and provides a "situational" dimension, where the CACL does not.

Current Reviews of Criminal Typologies

Megargee (1977) considered both the substance and form of a taxonomic system for offenders. He listed seven criteria for "usefulness" (p. 108) of such a classification scheme which are as follows:

1. The system should classify all of the offenders under consideration.
2. It should have clear operational definitions of types.
3. It should be reliable, especially across raters.
4. It should be valid (construct validity is implied).
5. It should be dynamic, reflecting changes in the individual.
6. It should carry implications for treatment.
7. It should be economical to administer.

These criteria do not stress the "theory building" function of such a system stressed by Sokal (1974), Schafer (1968) and others, but rather emphasize the practical significance of the system. In this, Megargee is following the point of view expoused by Gibbons (1975) in moving away from theoretically oriented taxonomic systems.

Commenting on the CACL diagnostic system, Megargee said: "Systems (such as the CACL) can reflect changes in the individual and typically have clear implications for differential treatment strategies" (p. 110). Later, he stresses the training and supervision of raters necessary to the CACL system, and said that the development of his MMPI system was intended to "retain the advantages of the Quay . . . system [and to be], . . . widely implemented with less cost and fewer trained personnel" (p. 110). However, he is equating a system based on a self-report device intended for⁷ psychiatric classification with a more direct system of behavioral monitoring. Thus, the comparison does not seem adequate in its inference that both are based on "personality characteristics of the offender" (p. 112), except in the broadest sense.

Examples of various typologies have been also discussed in a review of Warren (1969) and in the proceedings of an NIMH conference on criminal typologies (1967). Both of these reviews group typologies differently than does Schafer and elaborate other characteristics than those which he emphasized.

Warren discussed five groups of offender typologies which provide the background for her own classification system (p. 241). These typologies include the following:

1. Prior probability systems, which rank offenders on the expectancy of some future behavior, usually recidivism.

2. Reference group typologies, relating criminal behavior to the social norms of a specific group.

3. Behavior classifications, which are oriented to some aspect of the offender's behavior.

4. Psychiatrically-oriented approaches which seek to define the nature of any mental disorder underlying crime.

5. Social perception and interaction systems. Such typologies relate criminal behavior to specific social interactions, and to the criminal's perceptions of those interactions.

It is obvious that these groupings are poorly defined and that they frequently overlap, as Warren has admitted (p. 241). The reviewer continued to make several more valid points about the structure and function of offender typologies. Generally, Warren made the point that "each of the . . . classification systems is not equally relevant for all purposes" (p. 242).

Warren saw typologies as serving the purposes of either "management" or "treatment" (p. 242). She said,

It is possible for certain purposes to use a classification system which . . . has no etiological reference, one which has no implications for treatment, or one which is specific to an institutional setting. (p. 243)

Her review, like others, pointed out the difficulties with typologies which emphasize etiological dimensions, and argued for the use of more effective systems for treatment. In Warren's view, any combination of several factors may have caused the crime and it is necessary to specify the exact cause in order to change the behavior (p. 243). This view, although widely shared, has not led to more effective treatment of criminals. Warren reviewed several studies which indicated that no form of differential treatment has effectively reduced recidivism rates (p. 245).

Despite this fact, Warren remained optimistic that adequate typologies will reveal that treatment outcomes depend on characteristics of the offender which interact with characteristics of the treatment program. It is not surprising that her interpersonal maturity system is oriented to such a purpose. Unfortunately, no later articles have been published which report on whether her system was more effective than others.

Warren (1969) summarized the results of an NIMH study (1967) which attempted a cross-tabulation of many existing classification systems, including that of Quay. The resulting configuration of typologies or composite system revealed six "bands" which were judged to represent a

stable set of underlying characteristics of offenders (Warren, 1969, p. 249). The six categories common to the sixteen classification systems reviewed are as follows:

1. Band 1 - labeled the asocial type, included the CACL Psychopathic-Aggressive type. Such individuals are characterized as "primitive, underinhibited, impulsive, hostile, insecure, inadequate, maladaptive, demanding of immediate gratification and attention, thoroughly egocentric, etc." (Warren, p. 251).

2. Band 2 - labeled the conformist type, incorporated the CACL Immature-Dependent type. Persons in this band are characterized as "concerned with power, searching for structure, dominated by the need for social approval, rule-oriented, unable to empathize, having low, self-esteem" (Warren, p. 251).

3. Band 3 - labeled the antisocial manipulator, included the CACL Manipulative type. These offenders are described as "guilt-free, power-oriented, self-satisfied, non-trusting, emotionally insulated, cynical . . . and extremely hostile" (Warren, p. 252).

4. Band 4 - identified as the neurotic subtype, including the CACL Neurotic-Anxious type. Such individuals are characterized by high levels of anxiety and are described as "intimidated, disturbed, anxious, depressed, and withdrawn" (Warren, p. 254).

5. Band 5 - labelled as the subcultural identifier.

Such individuals are presumed to commit their crimes because of their integration of subcultural values conducive to crime. Individuals of this type are described as "loyal to their group, psychologically healthy, proud, adequate, suspicious of the authority system, having a stable family, have criminal attitudes, and accessible to new experiences" (Warren, p. 254).

6. Band 6 - labelled the situational offender. This grouping included the CALH situational type, and is characterized as "relatively normal, exposed to acute, severe stress, having no evidence of neurosis, having little prior criminal records, etc." (Warren, p. 255). Such persons are seen as reacting to an overwhelming, non-recurring emotional stress which led to committing their crime.

Unfortunately, these "bands" or subtypes were identified by an informal comparison rather than on the basis of the measurement of a heterogeneous group of offenders with the same group of classificatory instruments. That is, the bands were constructed intuitively rather than empirically. However, since the various classification systems were developed independently, it is possible that this consensus reveals the existence of separate constructs which differ across the bands. As Warren noted (p. 245), until an empirical study is done on a single population, the diagnostic bands described above will remain somewhat hypothetical and tentative.

Generally, Warren used this review to provide background for her own diagnostic system, but she made several points pertinent to criminal classification as it exists today. She pointed out that "the classification systems are not equally relevant for all purposes" (p. 241). In addition, this review indicated that an ideal typology would provide "an explanatory theory with the resulting aid to prediction, implications for management and treatment, greater precision for research" (p. 240). Thus it does seem that Warren believes that a single system can meet these needs.

As Gibbons (1975) pointed out, there is an increasing disenchantment with all of the taxonomic systems described above. Most have failed to show any real usefulness in the treatment of criminal behavior. Although the authors of these systems have hoped for empirical verification of their systems, little evidence has been forthcoming. Even though these systems have stimulated some new research, and do provide several of the benefits outlined by Sokal (1974), they have failed to show pragmatic usefulness (Schafer, 1968, p. 177).

Gibbons (1975) was also pessimistic about usefulness of current offender typologies. He said: "It is by no means clear that existing typologies are empirically precise" (p. 254). The reasons for this lack of clarity are several, according to Gibbons (p. 299). Firstly, no

single typology subsumes all types of criminality. Secondly, new forms of lawbreaking may be emerging which do not fit traditional typologies. Thirdly, the patterns of behavior or etiology which most typologies hypothesize have yet to be found in the actual study of offenders.

Gibbons argued that this lack of satisfactory classification systems is due either to the faults in the systems themselves or to the possibility that criminal behavior develops in a unique manner in each individual. It is difficult to assume the latter case, however, until the former has been eliminated as a potential problem.

He concluded by noting:

Insofar as the search for typologies turn out to be profitable in corrections, it will be as a consequence of the further development of statistical classifications . . . [which involve] . . . the development of classificatory devices based on specific groups of offenders within certain limited correctional settings. (p. 245)

Thus Gibbons recommended turning away from theoretically derived typologies, especially those which center on the etiology of criminal behavior. His discussion indicated that the more any typology depends on retrospective investigation or hypothetical constructs, the less likely it is to produce meaningful results.

In summary, the literature on classification systems for offenders seems to support several overall trends. First, no single system can perform all of the functions necessary in the criminal justice system. Monothetic,

crime-based systems are best suited to the needs of law enforcement agencies, while polythetic, treatment-oriented systems meet the needs of prison officials and program planners. Systems which have sought to delineate the causes of criminal behavior or to trace recurring patterns of adjustment prior to the offense have failed.

Second, since little empirical verification has been found for the theoretical constructs underlying many classification systems, the trend has been to attempt to define coherent sets of variables and to then explore the relations between these "categories" and other variables. That is, the usual process in developing an offender taxonomy has been to group individuals with similar crimes and then explore their similarities on other variables. As Megargee and Bohn noted (1977), this technique has been singularly unproductive, and the "psychometric" method which this system and the CACL use reverses this process. That is, offenders are categorized on variables related to current functioning, and the resulting types are related to past behavior or to predict future adjustment (p. 155).

Third, a tentative list of criteria for judging a taxonomic system for offenders emerges from all the articles in this area. These criteria are as follows:

1. It should relate to other variables of interest, and as a consequence may have theory building value.

2. It should serve a specific purpose for limited population.

3. It should assess current functioning rather than past behavior.

4. It should classify all individuals in question and no individual should be classified into more than one category.

5. It should have specific, clear-cut decision rules allocating individuals to categories.

This study will investigate the CACL primarily on related criteria (1) and (5). That is, the clearness of decision rules is reflected in the consistency of raters in assessing the behaviors in question. Thus the inter-rater reliability study will assess the clarity of the CACL's definitions, and the various validity studies will delineate the CACL's relationship with other variables. The following section will review the literature which provides the background for establishing the reliability and validity of the CACL.

Psychometric Concepts

This section reviews the concepts of reliability and validity as they pertain to this study. The classical theory of reliability and the major types of reliability estimates are reviewed. The factors affecting reliability and validity are then presented and the specific types of

reliability and validity estimates obtained in this study are discussed at greater length.

Reliability. The concept of reliability of measurement refers to its consistency across any of several dimensions. As several authors have pointed out, reliability, like validity, takes on special significance in the measurement of traits or inferred constructs (Stanley, 1969; Cureton, 1958). Reliability has been of central importance in areas such as psychology and education, where indirect measurement is frequently employed.

Definitions of reliability. Classical measurement theory has based the concept of reliability on the assumption that any measurement contains a discrete amount of random fluctuation or error in addition to the influence of the actual variable under consideration. As Stanley (1969) pointed out:

When a feature or attribute of anything (in any of the sciences) is measured, that measurement contains a certain amount of chance error. The amount of chance error may be large or small, but it is universally present. (p. 356)

Thus it is assumed that any observed score for an individual is composed of a true score component and an error score component which are linearly additive. Since the error score components are presumed to be random, they should not show any relationship to each other or to the true or observed scores. Cureton (1958) said:

The basic theorem which underlies all formulas of reliability, and of empirical validity as well, may be stated as follows: In a population of individuals, the errors of measurement in different tests and the different forms of the same test are uncorrelated with one another and are uncorrelated with the true scores on all tests and forms. (p. 103)

The error of measurement referred to by Cureton is an estimate which relates to the variability in a series of repeated testings of the same sample due to random (error) fluctuations. That is, if a number of independent measurements are taken on the same individuals, the variability in those measures would reflect the random fluctuations, or amount of error variance present in the measurements. The shared or common variance would reflect the amount of true score variability which was present. Since repeated testings of the same sample of individuals are not practical for a variety of reasons (the interactive effects of measurement, practice, etc.), the errors of measurement must be estimated indirectly.

It is possible to see that the notion of variability of scores across any of several dimensions is central to the definition of reliability. Without observed variance in scores, the estimation of reliability is not possible. Again, since the true score variance and error score variance can never be assessed directly, one must attempt to estimate them from the observed variance in test scores. As Stanley (1969) emphasized:

The basic problem in defining the reliability of a testing procedure . . . becomes that of defining what shall be thought of an error variance in relation to the type of inference one wishes to make from the test scores. When this definition has been made, the next step is to devise those series of empirical and statistical operations that will provide the best estimates of the defined fractions of variance. (p. 362)

Since variability in test scores can arise from a variety of sources, the selection of which of these are to be considered as sources of error variance depends on the purpose of the testing. Stanley emphasized this point when he said:

There is no single universal and absolute reliability coefficient for a test. . . . The allocation of variance from different sources calls for practical judgment of what use is to be made of the resulting statistical value. (p. 363)

The reliability coefficient for any measure can be defined as that proportion of observed score variance which is composed of true score variance. Theoretically, then, the reliability coefficient can range from 0 to 1.00, where a zero reliability coefficient indicates an absence of variability attributable to true score differences, and where a reliability coefficient of one results from a complete absence of random, extraneous variability. The formula for this relationship can be expressed as:

$$R_{tt} = \frac{S_t^2}{S_o^2}$$

where R_{tt} is the reliability coefficient, s_t^2 is the true score variance, and s_o^2 is the observed score variance. Since the observed score variance is presumed to be composed of a linear combination of true score variance and error variance ($s_o^2 = s_e^2 + s_t^2$), the formula may also be written as,

$$R_{tt} = \frac{s_o^2 - s_e^2}{s_o^2} \quad \text{or} \quad R_{tt} = 1 - \frac{s_e^2}{s_o^2}$$

Factors affecting R_{tt} . The presumed random nature and normal distribution of the error component influences the magnitude of the reliability coefficient in several ways. As the number of items in any measure increases, the errors will tend to cancel each other out to a greater degree. That is, as the number of items approaches infinity, the sum of the errors will tend to approach zero. Magnusson (1967) also noted that the error variance increases arithmetically with the length of the test while the true-score variance increases with the square of the number of items. Thus, "when the test is lengthened, the true variance increases at a faster rate than the error variance. This . . . means that the test will become more reliable" (p. 72).

In addition, the homogeneity or amount of total variance in the sample also determines the magnitude of the reliability estimate. As the sample becomes more homogeneous,

the amount of true score variance decreases, while the error remains unchanged. This decrease results in a reduction of the magnitude of the reliability coefficient, since the ratio of true score variance total score variance has been decreased.

Types of reliability estimates. As mentioned previously, the particular source of total test variance which is considered as error depends on intended use of the instrument. For example, if a measure is intended to measure a single, unitary trait it is highly desirable that the items share as much common variance as possible. Again, if a test is intended to measure an enduring characteristic of the individual, it should have as much stability across time as possible.

Reliability estimates can be thought of as approximations of true-to-total variance proportions, where the priority of the use of the test determines which of the above will be considered as most important sources of true variance. The various types of reliability coefficients can be thought of as falling into several broad classes, based on the type of error which is considered most important to the measure in question. Cronbach (1960) has defined three such classes of reliability coefficients which he calls coefficients of stability, coefficients of equivalence, and coefficients of internal consistency.

1. Coefficients of stability estimate the consistency of test scores across time, and are particularly important in measuring the lasting characteristics or traits of individuals. Such coefficients are generated in a test-retest paradigm where the same instrument is given on several occasions.

2. Coefficients of equivalence are intended to measure the similarity of several forms of a specific test. That is, equivalence estimates are intended to measure the degree to which two tests are parallel--that is, having the same means, variances, and average item intercorrelations. It is also possible to consider inter-rater reliability as a type of equivalence estimate, although the same form of the instrument is used. Inter-rater reliability estimates compare the shared variance across several individuals who assess the same person at the same time and under the same conditions.

3. Coefficients of internal consistency assess the degree to which the items in a test measure the same trait, construct or characteristic. One estimate of internal consistency is obtained by dividing the responses to a test into two parts, and correlating the two halves with each other, and approximating the reliability of the total test by the use of the Spearman-Brown prophecy formula. This is known as a "split-half" reliability coefficient.

Various other indices of internal consistency have been developed, such as Cronbach's coefficient alpha (Cronbach, 1951) and the Kuder-Richardson (1937) formulas, which estimate the average of all possible split-half reliability coefficients of a given test. These coefficients will not be further discussed, since this study is concerned only with the consistency across raters, rather than internal consistency or stability across time.

Reliability Estimation in This Study

The specific type of reliability which is of the greatest concern in this study is the degree to which equally trained independent observers agree on the presence of the behaviors assessed by the CACL. Although the instrument itself will be discussed further in the methodology section, it is important to note that it is neither a rating scale nor a traditional observational instrument. Rather than counting the frequency of occurrence of specific behaviors or rating the individual along a theoretical continuum, the CACL is designed to determine whether a specific behavioral trait is characteristic of the individual (Quay, 1964).

In a recent article, Frick and Semmel (1978) made several important points in regard to inter-observer agreement (reliability). They said:

Minimal observer disagreement is a necessary but insufficient condition for high reliability coefficients, since there are other components of the generic error variance that are theoretically independent from observer error variance (e.g., intrasubject variance from occasion to occasion). (p. 159)

In addition, the authors also note that although observer or rater agreement is only a part of the reliability of observational data, it does set the upper limit for the reliability of the data under consideration. That is, until the observational systems capacity for interobserver agreement has been defined, it is difficult to determine the degree to which other factors are limiting the reliability of the data (pp. 160-161).

Frick and Semmel also point out that the traditional definition of reliability as agreement between measures which have identical content, means, variance, and item intercorrelations is impractical when applied to human raters. That is, observers or raters do not have identical or equivalent observational skills. Accordingly, intraclass correlation coefficients or generalizability coefficients have been proposed as techniques to determine the reliability of a set of data without depending on the above assumptions. Such coefficients have often been used in the analysis of classroom observation data, but are equally applicable to measurements from other sources (Haggard, 1958; McGaw, Wardrop, & Burda, 1972).

Such coefficients estimate the ratio of true-to-total variance, but use an analysis of variance model to estimate the relative contributions of various sources of error variance. Although a more detailed description of the technique used in this study will be given in the procedures section, it is of importance to reiterate that such analytic techniques are used since the traditional assumptions underlying reliability are not applicable to data arising from ratings or observations.

An earlier article by Ebel (1951) compared the advantages of the intraclass correlation coefficient with other methods for assessing the reliability of ratings. In recommending the intraclass coefficient, Ebel listed three major advantages of such an approach.

First, the intraclass formula permits the investigator to choose whether to include "between raters" variance as part of the error variance. . . . Second, a convenient means for estimating the precision of the reliability coefficients is available to the user of the intraclass formula. Third, the intraclass formula uses the familiar statistics and routine computational procedures of analysis of variance. (p. 423)

In a position paper, McGaw et al. (1972) made a distinction between reliability coefficient as calculated from the internal structure of a test, from repeated testings, or from parallel forms, contrasting these with indices of observer agreement. Antedating the views of Frick and Semmel, they noted that agreement between observers has all

too often been considered the only important aspect of the reliability estimation of observational data. Specifically, they say:

The confusion introduced into the literature through failure to clearly distinguish the different sources of unreliability, and through over-emphasis on inter-judge agreement has resulted from a confusion of the importance of primacy with prime importance. Inter-judge agreement is the first, but not the most important issue to be faced. (p. 16)

Thus for the current study, it is most important to note that the inter-rater reliability (agreement) which is calculated is not to be considered the only aspect of stability of data arising from the CACL which should be studied. However, because of its importance it is the type of reliability to be examined in this study.

The inter-class correlation coefficients which were derived in the study are for the average of three raters, where each rater rates all subjects. These estimates are considerably higher than those which would be obtained for a single rater.

These coefficients are also calculated differently when absolute rather than comparative decisions are being made. When absolute decisions are involved, systematic rater bias is included in the error term of the model. For comparative decisions, such bias is not included along with the subjects by rater interaction in the error term.

Validity

Most authors agree that validity, like reliability, is a general term for a variety of related processes which assess the "usefulness" of a test. Brown (1970) pointed out that validity analysis may answer any of the following questions:

How well does the test do the job it is employed to do? What traits are being measured by the test? Is the test actually measuring what it was designed to measure? Does the test supply information that can be used in making decisions? What interpretation can be given to the scores on a test? What can be predicted from the test scores? (p. 99)

That is, validity studies generally attempt to relate test scores to other variables of interest. In terms of true and error score variance, Brown said:

Whereas reliability was defined by the proportions of true and error variance, validity is determined by the proportion of true variance that is relevant to the purposes of testing.
, (p. 98)

Thus, the process of validation usually involves assessing the relationship between the test and some external criterion.

The definitions of validity, which have been given in the Standards for Educational and Psychological Tests, center around the process of estimating the usefulness or meaningfulness of the data from a particular instrument. Each of these definitions will be discussed at a later point, but it is important here to compare the definitions of validity held by other authors.

Ebel (1961) suggested that defining validity is more difficult than it may appear at first glance. He pointed out that various authors diverge widely in their definitions of validity, and as examples notes that:

Gulliksen . . . has said: "The validity of a test is the correlation of the test with some criterion." Cureton writes: "The validity of a test is an estimate of the correlation between the raw test scores and the 'true' (that is perfectly reliable) criterion scores." Lindquist suggests: "The validity of a test . . . (is) . . . the accuracy with which it measures that which it is intended to measure. . . ." Edgerton suggests: "By validity we refer to the extent to which the measuring device is useful for a given purpose." Cronbach explains: "The more fully and confidently a test can be interpreted, the greater its validity." (p. 75)

Ebel continued by defining three other problem areas in the area of validity:

The fact that it must assume diverse forms to fit diverse situations, the discrepancy between the importance of test validity and the state of the art of validation, and the fact that the question of validity doesn't arise in the physical sciences. (pp. 76-78)

In addition, he pointed out that the concept of validity is not philosophically adequate, in that it is unlikely that, "the naive faith in the pre-existence of a quantity to be measured is basic to the general conception of validity" (p. 79).

Ebel also mentioned that these difficulties may well be due to a variety of causes. First, he suggested that although the relation between a test and criterion is central to validity theory, the criterion, like the test

itself is most often constructed and thus of limited validity itself. In addition to the philosophic problems of a "true" score, Ebel also saw the concept as frequently overgeneralized and used in inappropriate settings.

As a solution to these problems, Ebel (1961) suggests that the term "meaningfulness" be used to subsume the concept of validity. That is, he suggested that the assessment of the relationship between test scores and other measures be one of factors which contribute to the interpretability of test scores. He recommended the other factors to be considered should be the reliability of the measure, the norms used, and the operational definition of the score itself.

Following Ebel's recommendations, this study is an assessment of the meaningfulness of the CACL. That is, scores on the CACL are related to other measures for a sample which differs from the norms and the reliability of the instrument is assessed. In this way, we have an indication of the usefulness of the instrument with a population having a high degree of psychopathology.

Magnusson (1967) said that validity, like reliability, is an aspect of dependability, and that "the validity of a method is the accuracy with which meaningful and relevant measurements can be made with it" (p. 124).

As mentioned above, the criterion measure may be a test which has less than perfect validity and reliability

itself. Magnusson pointed out that although imperfect reliability can be corrected, "low validity in the criterion data, however, can never be corrected for . . ." (p. 127).

Often the question of how best to define the criterion variable is left essentially unanswered.

Types of Validity

Other authors concur that validity is most often concerned with the relationship between the test and other variables. Like reliability, this relationship can exist in any of several dimensions. Each of these dimensions covers a different aspect of validity, and may be thought of as the relationship between the test and a larger domain, other measures of the same trait, or the degree of "meaningfulness" of the test. The types of validity which correspond to those dimensions have been mentioned above and labelled by the American Psychological Association as content validity, criterion-related validity, and construct validity (APA, 1974).

The first of these concepts, content validity, refers to the adequacy with which a measure reflects the domain of items in question. Although content validity is an important area in the construction of achievement tests, it has little bearing on this study. Therefore, it will not be discussed at length.

Criterion related validity has been defined by Gaion (1974) as "the extent to which scores on one variable, usually a predictor, may be used to infer performance on a different and operationally independent variable called a criterion" (p. 288). If the criterion measure is taken at the same point in time, the process is known as concurrent validation. If the measure is taken later, the process is known as predictive validation.

As has been mentioned previously, validation studies are intended to specify the "usefulness" of the test, or the degree to which it successfully accomplishes a given purpose. In a general review of validation, Cronbach (1960) equated criterion related validity with usefulness in selection and placement, both of which he subsumes under the process of decision making (p. 446).

It is important to note that criterion-related validity may be conceptualized as existing for a specific purpose and is empirically determined by the relationship between the test scores in question and a second criterion measure. In a brief review, Cureton (1958) said that the criterion may exist in the present or future, and may be pre-existing or constructed (p. 105).

Pre-existing criteria include those that exist without any special effort made to predict them. Examples of such criteria include graduation from college, number of previous criminal convictions, etc. Constructed criteria are

usually developed on the basis of some hypothetical trait concept, and include rating scales, intelligence measures and personality tests.

Criterion-related validation studies often numerically express the relationship between this test score and external measures in the form of a validity index, which represents the amount of variance common to the two. However, it is often presumed that the criterion measure is an adequate measure of the criterion when in reality this may not be the case. In an article on the problems inherent in criterion-related validation, Brogden and Taylor (1950) defined "criterion bias" as "any variable, except errors of measurement and sampling errors, producing a deviation of obtained criterion scores from a hypothetical 'true' score criterion" (p. 82).

Although bias in the criterion which is not correlated with the predictor may undesirably affect validity studies of this type, Brogden and Taylor point out, "it is the presence of test-correlated bias that 'makes' or 'breaks' the criterion" (p. 82).

Construct Validity Estimates

Unlike criterion related validity, construct validation procedures are often more conceptual than statistical. They attempt to assess the degree to which an instrument reflects an underlying construct or hypothetical trait. In a classic article, Cronbach and Meehl (1955) stated:

Construct validation is involved whenever the test is to be interpreted as a measure of some attribute or quality which is not "operationally defined" Construct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured. (p. 282)

The authors continued to point out that construct validity is "not to be identified solely by the particular investigative procedures, but by the orientation of the investigator" (p. 281). That is, the procedure may incorporate concurrent or predictive methodologies, factor analysis, or other techniques to be discussed in this section. It is the aim or intent of the investigator that uniquely defines construct validation.

A number of procedures have been used in an effort to determine the usefulness of a given construct in interpreting test data. Cronbach and Meehl listed several such techniques which provide the basis for inferring the existence of a trait. These techniques include the following:

1. Studies of group differences which would be expected on the basis of the construct in question.
2. Correlations between items or tests which reflect the same trait. The covariation between such items or tests may be measured by means of factor analysis and correlation matrices.
3. Studies of the internal structure of the measure in question. For many constructs, evidence of homogeneity within the test is relevant in judging validity.

4. Studies of change over occasions (retest reliability) may lend support to the logical network defining the construct.
5. Studies of the process of performing on the measure in question may also help to define the construct in question. (p. 289)

In a reformulation of the techniques mentioned above, Campbell and Fiske (1959) point out that although we often use measures of association (correlation) to assess the presence of a construct, we also often look for divergences in test performance. They define the two processes as in the following manner:

1. Validation is typically convergent, a confirmation by independent measuring procedures. Independence of methods is a common denominator among major types of validity (excepting content validity) insofar as they are to be distinguished from reliability.
2. For the justification of novel trait measures, for the validation of test interpretation, or for the establishment of construct validity, divergent validation as well as convergent validation is required. Tests can be invalidated by too high correlations with other tests from which they were intended to differ. (p. 82)

That is, the process incorporating convergent and divergent validation indices aids specifically in the logical interpretation of validation data. By demonstrating that different techniques intended to measure the same trait correlate significantly with each other, and that similar methods intended to measure different traits do not, powerful logical evidence for the traits' presence has been presented.

Following Campbell and Fiske's logic, it is evident that construct validation relies on both statistical and logical inferential techniques. That is, it uses empirical evidence to logically deduce the presence or absence of a specific trait. Unlike criterion-related validity, which relies heavily on statistical measures of association, the construct validity of an instrument is demonstrated through a series of analyses which are logically incorporated into the overall validation process.

Factor analysis is widely used in the determination of construct validity. An early article by Guilford (1948) stressed the use of factor analysis in assessing the construct validity of an instrument. Guilford seemed to be anticipating the distinction between criterion-related and construct validity when he wrote of practical and factorial validity. He defined the factorial validity of a test as being determined by "its loadings on meaningful, common, reference factors" (p. 428).

Cattell (1964) also discussed the use of factor analysis in the determination of construct validity. As a type of convergent validity, he believed that factor analysis can help to define a construct when it emerges as a simple factor across several studies. This technique "combines measurement precision with unitary character, as well as a meaning enriched beyond that of an empirical construct" (p. 22).

Although Anastasi (1976) indirectly accepted the use of factor analysis in construct validation, particularly with reference to the measurement of general versus specific abilities, her overall stance has been strongly against anything other than criterion-related validity. She referred to "the will-o'-the-wisp" of psychological processes which are distinct from performance" (p. 77). Cronbach and Meehl (1955) disagree with this position, and point out that inference based on patterns of association between variables "cannot be dismissed as pure speculation" (p. 290).

The CACL was not developed to measure a prespecified underlying trait, but rather was developed through the factor analysis of a set of behavioral descriptors. However, the four subscales of the CACL have been given labels based on their content, and these have been shown to correspond to broader traits which have appeared throughout various classification systems for offenders. Thus, the validation process in this study will attempt to relate scores on the subscales of the CACL to other measures which may be indicative of those traits. In this sense, estimates of construct validity are of primary importance in this study. That is, it is most important to define the nature of traits measured by the instrument, rather than to only establish its estimates of criterion-related validity.

Chapter Summary

Included in this chapter is a review of the literature in three major areas which are pertinent to this study.

First, the process of classification in general has been summarized. Generally, classification systems serve many purposes and no single system can meet all the needs in any one area. Next, in reviewing the history of classification in the field of criminology, the problems with theoretically oriented typologies have been noted. Empirically derived classification systems such as the CACL after the advantage of proven utility for a specific population but need to be reevaluated before they are used with a group which differs from the normative sample.

Since the purpose of this study is to evaluate the psychometric proportion of the CACL based on the behavior of a group of mentally disordered criminals, the area of reliability and validity were reviewed at some length in this chapter. Particular emphasis is given to the topic of inter-rater reliability, which sets the upper line for the reliability of rating scale such as the CACL.

Criterion-related validity was also discussed at some length since the CACL is intended to facilitate decisions about future custody and treatment of individuals in confinement. This study relates the CACL to several criterion measures, including the MMPI and behavioral measures of disruptiveness.

Since these behavioral measures are of interest in their relation to the hypothetical traits measured by the CACL, the area of construct validity is also reviewed. Although the CACL is designed to describe patterns of behavior within the institution, it also labels these patterns in accordance with existing theories of criminal behavior. Thus, it may be used in "theory-building" studies rather than as a descriptive tool.

.

CHAPTER III

METHOD

The purpose of this study was to investigate the psychometric properties of the Correctional Adjustment Checklist (CACL), based on ratings of the behavior of a group of individuals confined in a maximum security mental hospital. Specifically, this study was designed to assess the inter-rater reliability of the instrument and to provide estimates of its construct and criterion-related validity when used with individuals showing evidence of various types of mental disorders.

The procedures used to obtain these estimates are detailed in a description of the subjects, the instruments, and the analytic techniques used. Since the emphasis of this study was to evaluate the instrument when used with a group which is different from the normative sample, the description of the subjects which follows is of considerable importance.

The Sample

All subjects included in this study were housed in the North Florida Evaluation and Treatment Center (NFETC), which is a 225-bed maximum security mental hospital located in Gainesville, Florida. It is operated and administered

by the Department of Health and Rehabilitative Services of the State of Florida and is currently the only mental hospital in the state which serves a purely forensic population.

The hospital is composed of eleven residential and treatment buildings, consisting of one to three nine-person living areas which are known as "pods." Each patient (known as a resident) has a private room, and shares bathing facilities and a living area with the other residents in his pod. The hospital is divided into three units, each of which serves a particular type of client. Based on diagnostic categories, these types are as follows: psychotic, behaviorally disordered, or mentally disordered sex offenders.

Although all of the residents have been charged and arrested for a major felony, not all have been tried, convicted, or sentenced. Those individuals who have been found incompetent to stand trial or to be sentenced are placed in the psychotic unit for short-term (averaging two months) treatment. Also, individuals who become psychotic while incarcerated are given similar short-term care. The Psychotic Unit currently includes ninety beds.

The Behavior Disorders Unit is comprised of forty-five beds and is intended for the behavioral management and treatment of antisocial, retarded, or neurologically impaired individuals. Such persons are usually management

problems in the traditional prison system, and are sent to NFETC for short-term treatment of recurring problem behaviors.

The Sex Offender treatment unit includes ninety beds and is oriented to the long-term (approximately two years) treatment of individuals who have been convicted of a sexual offense and been classified under Florida Statute 917 as Mentally Disordered Sex Offenders. The individuals so classified must be manifestly non-psychotic, and be judged by at least two psychiatrists to have a predisposition to commit other sexual offenses.

Overall, the population of the North Florida Evaluation and Treatment Center can be described as a group of approximately 225 males, all of whom have been arrested for a major felony and most of whom have been either found incompetent to stand trial or incompetent to be sentenced; who have become psychotic or a management problem while incarcerated; or who have been adjudicated as Mentally Disordered Sex Offenders. The age of the residents at the time of this study ranged from seventeen to seventy-nine, with a median age of twenty-eight, and they came from a wide variety of ethnic and social backgrounds within the state of Florida.

Selection of the Sample

From October 1976, when NFETC first began receiving residents, until July 1, 1978, approximately 550 individuals

have been treated or evaluated in the institution. Of these, approximately 325 have been treated and returned to the referring agency, while the remainder are still confined at the hospital.

The data which are available on these individuals are a function of events which were not under the control of this author. Since the emphasis at this hospital is on treatment and effective management of residents, changes in intake and diagnostic procedures were made which did not allow data collection procedures which would have been optimal for this study.

For the first 14 months of operation (until January 1978), the hospital included a central intake and diagnostic unit where all incoming residents were housed for short-term evaluation and diagnosis. During their stay in the intake and diagnostic unit, the residents were assessed on a battery of diagnostic tests including the Minnesota Multiphasic Personality Inventory (MMPI), the Incomplete Sentences Test, the Social Reaction Inventory, the Quay Correctional Adjustment Checklist (CACL) and Checklist for the analysis of Life History (CALH).

Since January 1978, the Intake and Diagnostic Unit has been concerned with the evaluation of incoming sex offenders only. Admission of residents to the Psychotic and Behavior Disorders Units has been directly to the building in which they were to be treated. This change has occurred

because of increased number of admissions to the Sex Offender Unit and because of the increased need for more intensive evaluation of incoming residents.

Accordingly, the Intake and Diagnostic Unit has increased the number of evaluation instruments which are administered to sex offenders. All sex offenders are given the MMPI, CACL, Bipolar Psychological Inventory, a short form of the Wechsler Adult Intelligence Scale, the California Psychological Inventory (CPI), and a complete and extensive social and demographic background information survey. Descriptive statistics for this sample are presented in Table 1.

Thus, most of the residents who have been admitted to NFETC have been tested during the first week of their stay in the institution. Unfortunately, since January of 1978, many residents who have been admitted to the Psychotic and Behavior Disorders Units have not been rated on the CACL. Since the residents were admitted directly into treatment, the staff in the buildings in which they were placed had not been trained in the use of the CACL or other diagnostic instruments.

Accordingly, the sample on which the following study of the CACL is based includes higher proportions of Mentally Disordered Sex Offenders than other treatment categories. Although some test data are available on all residents, with few exceptions, only those who were rated on the CACL during the first two weeks of their stay at NFETC are

included in this study. The exceptions to this sampling plan are those 27 individuals who were included in the inter-rater reliability study. Those persons had all been in treatment in the Sex Offender Unit for at least 60 days.

TABLE 1
NUMBER OF RESIDENTS BY UNIT ADMITTED TO NFETC
FROM ITS INCEPTION UNTIL JULY 1, 1978

Psychotic Unit	90	179 .
Sex Offender Unit	90	45
Behavior Disorders Unit	45	91
	In treatment as of 7/1/78	Discharged prior to 7/1/78

Of those residents admitted, intake data on the CACL are available on 140 individuals. Of these, 73 have been treated in the Psychotic Unit, 47 in the Sex Offender Unit, and 20 in the Behavior Disorders Unit.

The number of residents included in each of the studies reported here varies to some degree as a function of the availability of CACL intake data. While the central Intake and Diagnostic unit was using the CACL, each resident was rated independently by three staff members, and an average

rating was used to describe the individual. The reliability of the ratings on the 140 individuals on whom such data are available will be computed and compared with that obtained on the twenty-seven residents who were included in the sex offender sample.

After January 1978, the CACL was administered only to those residents who were considered diagnostic problems or whose placement in a particular treatment unit was difficult. All residents were given the MMPI within two weeks of the date they entered NFETC, and often were retested if their responses were considered invalid. If this is the case, the second profile is used for the studies described here.

Instrumentation

The primary instrument of interest in this study is the Quay Correctional Adjustment Checklist (CACL). This is a 41-item, factor analytically derived behavioral checklist. It was developed between 1964 and 1971 as a classification instrument for incarcerated males. In form, it is neither a true rating scale nor behavioral checklist. Rather, it includes a number of statements which are said to be characteristic of the individual in question.

The CACL is related to the early work of Hewitt and Jenkins (1946) who conducted an analysis of clusters of traits common to juvenile delinquents referred to a child guidance clinic. The resulting groups of traits were used

to classify juvenile offenders into three categories: unsocialized-aggressive, socialized delinquent, and over-inhibited.

Based on these results, Quay (1964) developed a 36-item checklist which was used to quantify the life histories of approximately 100 juvenile offenders. The responses to this checklist were factor analyzed in order to determine whether patterns of developmental events could be used to classify juvenile offenders. The results of this study indicated that the categories developed by Hewitt and Jenkins also appeared in the data obtained by Quay (1964). The checklist itself was later developed into the Checklist for the Analysis of Life Histories (CALH), which is often used as a supplement to the CACL.

Subsequently, Quay reported on the development of the CACL and CALH in a 1971 paper. In describing the development of the CACL, Quay related that a pool of behavioral descriptors was assembled from correctional workers and from previous research. Approximately 1,000 inmates from four institutions were rated on the items which were derived from these traits, and the resulting data were analyzed by means of factor analysis in order to estimate the extent of any underlying traits in these results. Four factors emerged, three of which correspond to those found in the CALH.

In describing the item selection technique used, Quay said that analyses were performed on three separate samples, each drawn from a different Federal Correctional Institution. He noted that,

Subsequent to the first analysis, items which did not meet the frequency criterion (not more than 90% or less than 10% of the subjects were rated as exhibiting the trait) and items which loaded less than .20 on any of the factors were dropped, and other items were added for the second analysis. (Quay, 1971, p. 3)

All three analyses produced four principal dimensions. The first, labeled Aggressive-Psychopathic, reflects toughness, defiance, physical and verbal aggression, trouble-making, victimizing, and quick temperedness. The second dimension, labeled Immature-Dependent, is composed of such behaviors as inability to follow directions, sluggishness, daydreaming, preoccupation, passivity, moodiness, and dullness. The third factor, given the label Neurotic-Anxious, reflects worry, tenseness, help seeking, fear of other inmates, sadness and emotional lability. The fourth dimension, measured by only five items is labeled as Manipulative and involves such characteristics as trying to "con" staff, lack of trust of staff, accusing staff of unfairness, and playing staff against one another.

According to Quay, the factors which emerged in the three samples were congruent with each other to a high degree in two cases (the Psychopathic-Aggressive and Immature-Dependent subscales), and less so in the cases of

the Neurotic-Anxious and Manipulative subscales. The degree of congruence was measured by Tucker's congruency coefficients, but the numerical values of these coefficients were not presented by Quay.

In the final selection of items, two major criteria were used: first, the item had to have a loading of .40 or higher in one or more of the analyses described previously; and second, the item had to load on the same factor in two of the three samples. After items were selected on these criteria, the results from all three groups were combined and factor scores were computed using unit weights. That is, each item checked as characteristic of the individual earned a value of one toward the score on that factor. Thus, the maximum score on each factor is the number of items contained on that subscale.

When the raw score distributions for each scale were plotted, Quay reported "gross departures from normality" (p. 5), which were evident by visual inspection. The raw scores were subsequently converted to normalized "T" scores.

As an estimate of the internal consistency reliability of the CACL, Quay reported that alpha coefficient was calculated for each of the subscales. For the total sample of 829 (all three groups combined), the reliability estimates were as follows: .91 for the Psychopathic-Aggressive subscale; .82 for the Immature-Dependent subscale; .77 for the Neurotic-Anxious subscale, and .77 for the Manipulative subscale.

Quay also examined the intercorrelations of the four subscales. He noted that: "While the factor analytic procedure results in uncorrelated factors, the actual estimates of scores of individuals on the factors are not necessarily independent" (p. 5). The highest intercorrelation (.81) was found between subscales 1 and 4 (Psychopathic-Aggressive and Manipulative). A moderate correlation (.41) was also found between the Immature-Dependent and Neurotic-Anxious subscales. Quay speculates that this is probably due to rater's tendency to evaluate prisoners as being "totally troublesome" (Quay, 1971, p. 5).

Quay (1971) also reported a validation study in which CACL subscale scores were related to a variety of other variables, primarily demographic in nature. He reported that all of the subscales showed a "modest" relationship to other variables. The Psychopathic-Aggressive subscale correlated negatively with the age of the criminal and positively with the number of prior arrests. The Immature-Dependent subscale tended to relate negatively to I.Q. and years of education. Scores on the Manipulative subscale tended to relate negatively to number of prior arrests, but exact numerical values were not presented.

In general, the CACL has fairly high internal consistency, but unknown inter-rater reliability. Although it was designed to provide subscales which are independent of each other, modest subscale correlations are found in most

studies. Evidence of construct validity is slight; statistically significant correlations exist between CACL subscales and some other variables, particularly number of prior arrests, age at arrest, and intellectual level.

No specific suggestions for decision rules are included with the instrument, forcing the user to choose whether to use scores on the CACL in making absolute or comparative decisions. When the instrument was used at NFETC, the highest subscale "T" score determined an individual's CACL classification type. This classification was supplemented by other tests, interviews and so forth.

Other instruments have been used to classify individuals who are incarcerated. Such instruments range from projective tests such as the Rorschach Ink Blots to self-report inventories such as the 16 Personality Factor Inventory. It is of interest that these instruments were not created for the purpose of classifying criminals, but were developed as diagnostic aids in mental health settings.

The Minnesota Multiphasic Personality Inventory (MMPI) is a 556-item self-report personality inventory. It was developed as an aid to the classification of psychiatric patients, and each of its original nine subscales corresponds to a diagnostic category current at the time of the test's construction. Although these categories were originally presumed to be mutually exclusive, subsequent research has shown this not to be the case (Dahlstrom, 1972).

Despite the intercorrelation of the subscales as well as the tests' sensitivity to the response set of the test-taker (Messick & Jackson, 1967), the MMPI has been shown to be useful in assessing a variety of areas of functioning. Recent research has stressed the interpretation of profiles of subscale scores rather than classification into one of several psychiatric diagnostic categories (Meehl, 1955).

In addition to the original nine diagnostic subscales, three "validity" subscales were added to the instrument. These scales are intended to estimate the interpretability of the other subscales, and measure ego strength, naive lying to "fake good" and the frequency of items seldom endorsed by the normative population.

In general, the MMPI has been shown to be more valid for whites than blacks and to discriminate accurately between groups of psychiatric patients and prisoners with accuracy. Local norms are often more useful for behavioral predictions than are national norms (Palmer, 1970), but both provide predictive ability at a level significantly above chance. The MMPI has also been shown to relate to several other measures of criminal behavior, both in and out of incarceration (Panton, 1966).

By estimating the nature and extent of the relationship between the MMPI and the CACL, it is possible to assess the traits common to both and the overlap or redundancy in the instruments. It is also possible to estimate

the relationship between the CACL and other variables which are of interest in themselves, as well as for their logical relation to the traits measured by the CACL. Disruptive behaviors in the institution constitute such a criterion variable transition.

For the purpose of this study, disruptive behavior was defined as any act which was contrary to the resident rules of the North Florida Evaluation and Treatment Center and which disturbed the ongoing course of treatment. Such behaviors usually necessitate staff intervention and were limited to threats of aggression, aggressive acts, threats of self-injury, acts of self-injury, destruction of property, and other unclassified infractions of rules (violation of curfew, refusal to take medication, etc.)

Data Collection

The data used in this study were collected at different times, by different staff members at NFECT, and on different individuals. As part of the normal intake procedure, ratings on the CACL as well as scores on the MMPI were obtained on 140 residents. In addition, CACL ratings were also obtained on a smaller, more homogeneous group of individuals who had been observed for at least eight weeks. Finally, the frequencies of six types of disruptive acts were recorded and included as a behavioral

adjustment to confinement. CACL intake data were collected from October 1, 1977, until July 1978, as were MMPI scores on the same individuals. Ratings on the CACL for the smaller group of residents were collected in May 1978.

To obtain the measures of behavioral adjustment, a tally of disruptive behaviors was made from the daily observation notes kept on each resident. These notes were written by the treatment staff in each building at least once every eight-hour shift. Since all significant behaviors, especially infractions of rules, were to be included in observation notes, it seems likely that most disruptive behaviors were so recorded.

For each of the 140 residents on whom CACL and MMPI data were available, a survey was made of the 180 observation notes written during the first 60 days of his confinement. Those individuals who stayed less than 60 days were not included in this section of the study. Each note was inspected to determine if any disruptive behaviors were recorded. If more than one such behavior was mentioned, each was tallied separately. That is, if a resident threatened a staff member after receiving an infraction for face count, two disruptive behaviors were tallied. Only the specific mention of behaviors observed directly by staff were included. If one resident informed on another, the disruptive behavior was not tallied unless it was directly witnessed by a staff member.

For each resident included in this study, a record of the most recent arrest and conviction was made from the FBI "rap sheet." This is a listing of all prior arrests and convictions for the individual in question and is compiled from all arrest records throughout the United States. Arrests and convictions are matched by fingerprints as well as by name, so that crimes committed under an alias are also included. For this study, the following categories were used: murder, armed robbery, assault (including attempted murder), breaking and entering, forgery, and other nonviolent property crimes, rape or sexual assault, and nonviolent child molestation.

Although the first two analyses included in this study both assess the inter-rater reliability of the CACL, they differ in several respects. The first is based on the ratings made by three staff members of the intake and diagnostic unit. They were made after a relatively short (seven-day) period, which according to Quay (personal communication, 1978) may not allow for sufficient observation time. The second study is based on ratings made by three staff members in the sex offender treatment program. These ratings were based on the behavior of 27 residents who were being treated in that program, and who had been in treatment for at least eight weeks. The raters had observed the residents for the duration of their stay in treatment, and thus had the opportunity to

base their ratings on a larger sample of behavior than that in the first study. In both studies, each rater rated every subject.

The raters for the second study were trained over a seven-day period. Their training included operational definitions of the behaviors assessed by the CACL, as well as comparisons of their ratings on the same residents. That is, the raters filled out a CACL on two residents without discussing the results with each other. These ratings were then compared on an item-by-item basis, with group discussion of any discrepancies. After three such sessions, the raters agreed on 90% of the items on the CACL, and training was discontinued.

Data Analysis

All data were analyzed at the University of Miami computing facility using a UNIVAC 1100 computer. All analyses requiring a "packaged" computer program used the Statistical Package for the Social Sciences (SPSS) which is available in several versions at the University of Miami.

Reliability of the CACL

The inter-rater reliability of the CACL was estimated by the use of the intraclass correlation coefficient, which has been described by Ebel (1951) as well as by Bartko (1966). Essentially, this method uses the analysis of variance to estimate the proportion of variance in a set

of measurements which can be attributed to individuals, raters, and error. In this case, a subject by rater design was used. The resulting mean squares from the analysis of variance were substituted into Ebel's (1951) formula. As expressed by Ebel, that formula is:

$$v_1 = \frac{M_{\bar{X}} - M}{M_{\bar{X}} + (k-1)M}$$

The v is the intraclass correlation coefficient, $M_{\bar{X}}$ is the mean square for individuals, M is the mean square for error and k is the number of observers or raters. This formula is for estimating the reliability of a single rater, and does not include systematic rater bias in the error term.

When used in making absolute decisions, any systematic bias of the raters needs to be included in the error term of the formula. Thus, the formula is:

$$v_2 = \frac{M_{\bar{X}} - M}{M_{\bar{X}} + (k-1)M + k(M_{\bar{R}} - M)/N}$$

In this case, v_2 is the intraclass correlation coefficient, $M_{\bar{X}}$ is the mean square for subjects, M is the residual mean square, k is the member of raters, $M_{\bar{R}}$ is the mean square for raters, and N is the number of subjects.

Since the average of three raters scores was used in placement decisions at NFETC, a third formula was used

to provide estimates of the reliability of the average. Including systematic rated bias, that formula is:

$$v_3 = \frac{M_{\bar{x}} - M}{M_{\bar{x}} + k(M_{\bar{R}} - M)/N}$$

If we exclude systematic rater bias, the formula for the reliability of the average of 3 raters becomes:

$$v_4 = \frac{M_{\bar{x}} - M}{M}$$

This formula was used to estimate the reliability of the sum of all four subtests, as well as for each individual subtest. It should be noted that one sample on which these observations were drawn was fairly homogeneous, in that it consisted of only one type of offenders; i.e., Mentally Disordered Sex Offenders. This homogeneity may have provided reliability estimates which are somewhat less than maximal. The other inter-rater reliability study was based on the ratings of all types of incoming residents over a nine-month interval. The training of the raters was not under the control of this author, and residents were rated after a relatively short period of observation.

Predictive Validity of the CACL

The ability of the CACL's subscale scores to predict institutional disruptiveness was estimated through a

multiple regression procedure. This technique analyzes "the collective and separate contributions of two or more independent variables . . . to the variation of a dependent variable" (Kerlinger & Pedhazur, 1973, p. 3). That is, it estimates degree of relationship between a set of two or more variables and a single other-variable of interest. It provides approximations of the contributions to the variance of the dependent variable by a group of independent variables. This is accomplished by minimizing the sum of squared deviations between the predicted dependent variable values and those actual values obtained in the experiment. A linear combination is derived for the independent variables which minimizes those errors of prediction.

The model for this "least squares" solution can be expressed as:

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + E,$$

where B_0 is a constant value, and $B_1 \dots B_k$ are the weights assigned to the independent variables $X_1 \dots X_k$.

The weights in a regression equation which are based on the raw scores of the independent variables are known as partial regression coefficients. They are scale dependent, in that they are not directly comparable with each other in absolute magnitude. These weights may be transformed into standard score format so that they are

directly comparable in size. In this case, the weights are known as standardized partial regression coefficients, and reflect the unique contribution of each independent variable to the variance explained in the dependent variable.

In this study, the total frequency of disruptive behavior during the first 60 days of incarceration was the dependent variable, and the subtests of the CACL were the independent variables entered in the multiple regression equation. The subscales of the MMPI were also entered in a separate analysis in order to compare the predictive validity of the two instruments.

Construct Validation of the CACL

The construct validation of the CACL was carried out by means of a canonical correlation analysis. This technique, described by Timm (1975) and others, is an extension of the multiple regression method discussed previously. That is, it provides an estimate of the maximum correlation possible between two linear composites of two sets of variables. This is accomplished by including more than one dependent variable in a linear composite which maximizes the degree of relationship between that group and a linear composite of independent variables.

Two sets of predicted scores are generated by these linear combinations. If the dependent variables are identified as $\hat{y}_1 \dots \hat{y}_n$, then $\hat{u} = a_1\hat{y}_1 + a_2\hat{y}_2 + \dots + a_n\hat{y}_n$,

where \hat{u} is the composite value. If the independent variables are labeled as $x_1 + \dots + x_n$, then $\hat{v} = b_1\hat{x}_1 + b_n\hat{x}_n$, where \hat{v} is the composite of those values. Thus, a canonical correlation analysis provides the weights $a_1 \dots a_n$ and $b_1 \dots b_n$ such that the Pearson Product-Moment correlation between \hat{u} and \hat{v} is a maximum.

In this study, canonical correlation analysis was performed to relate the subscale scores on the CACL to subscale scores on the MMPI, with data on both instruments being collected at the time of intake. The results of the canonical correlation analysis were used in a redundancy analysis, which has been described by Stewart and Love (1968). This technique provides a numerical estimate of the redundancy in one set of data, given the other. The redundancy coefficients were obtained by rating the proportion of variance in each set of variables extracted by each canonical variate. These proportions were then multiplied by the corresponding squared canonical correlation and summed across the significant canonical variates for each set separately. The resulting coefficients represent the proportion of variance in a set of variables that may be explained by the second set.

Postdiction of Crime Type

In order to further determine the validity of the CACL, a discriminant function analysis was performed using the CACL subscale scores as independent variables and type of crime as a categorical dependent variable.

A discriminant function analysis is an extension of the multiple regression procedure, in that it provides a set of weights for the independent variables which minimize the errors of prediction when the dependent variable is group membership. As in multiple regression, a linear combination of independent variables is formed such that $y = a_0 + a_1y_1 + a_2y_2 \dots a_ny_n$, where $a_1 \dots a_n$ are the weights for the independent variables $y_1 \dots y_n$. This linear combination provides the best discrimination between the groups by maximizing the among group variance in relation to the within group variance.

In this study, only the most recent conviction was used to determine crime type. Nonviolent crimes were defined primarily as crimes against property (breaking and entering, etc.), while violent crimes were defined as those which involved physical aggression toward another individual (rape, assault and battery, homicide, etc.). The subtests of the CACL were used as the independent variables in the equation.

Summary

In this chapter, the sample and instruments used in this study are described as well as the procedures for data collection. It is noted that the data collection procedures were not under the direct control of this investigator and thus introduce certain limitations.

Also included in this chapter is a description of the various procedures used in the analysis.

Separate analyses were conducted to obtain reliability and validity estimates of the CACL. Two estimates of inter-rater reliabilities were obtained on each subscale of the CACL. The first set of inter-rater reliabilities was computed using the ratings from observers who were not trained on a sample of incoming residents to the institution. Intraclass correlation coefficients were computed from the data. The second set of inter-rater reliabilities used the same method of computation on ratings by trained observers on a sample of sex offenders.

The relationship between the CACL and the MMPI was assessed using canonical variate analysis. Canonical correlations and redundancy indices were computed. In addition, multiple regression analysis was used in the prediction of institutional disruptiveness from the CACL. Finally, a discriminant function analysis was used to predict type of crime based on the CACL subscale scores.

CHAPTER IV

RESULTS

The results of the analyses described previously are presented in this chapter. Generally, the results are given without interpretation, since their explanation and synthesis are presented in the following chapter. Descriptive statistics precede each section of this chapter.

In the first section of this chapter, the results of the two inter-rater reliability studies are presented. The first presents the reliability estimates for the ratings done on intake (after 4-7 days of observation) by raters whose training was not controlled by this writer. The second provides a summary of the "controlled" study in which the raters had been trained by this writer and where the subjects had been observed for a minimum of thirty days.

The second section of this chapter contains the results of the canonical correlation analysis between the average ratings on the CACL at intake and the scores on the MMPI administered at the same time. This section is followed by a presentation of the correlations between the canonical variates and the original variables to clarify the content of the canonical variates. The results of the redundancy analysis are also included.

The third section includes the results of the series of multiple regression analyses relating scores on the CACL to several types of disruptive behavior. Results are given separately for suicide attempts, assaultive behavior, verbal threats and coercion, as well as for other infractions of program rules. This was done in order to relate CACL subscale scores to specific types of disruptive behavior, in order to assess the relationship between those behaviors and the CACL subtest which would be expected to relate most strongly to them.

The final section of this chapter contains the results of the discriminant function analysis, which relates scores in the CACL subscales to the presence of violence in the crime for which the subject had been most recently arrested and/or convicted. That is, scores on the CACL are weighted so that a linear combination of subtests best predicts group membership, where the criterion for group membership is the presence or absence of violence.

Inter-Rater Reliability of the CACL

As has been explained previously, two studies of the inter-rater reliability of the CACL were performed. These studies used separate samples of subjects and raters, and the reliability estimate from each was obtained through an analysis of variance procedure. The descriptive statistics for the "intake" and "controlled" studies are presented in

Tables 14 and 15 in Appendix B, respectively; Tables 16 through 23 include the corresponding analysis of variance summary tables for each subtest.

Separate analyses were performed for each of the four subtests of the CACL for the average of three raters, as well as for single raters. These estimates are given for the average rating first, followed by that for single raters. For the "controlled" condition the coefficients are: I-D Subscale, $r=.37$, $.26$; P-A Subscale, $r=.76$, $.51$; N-A Subscale, $r=.73$, $.46$; and Ma Subscale, $r=.78$, $.59$. For the "intake" condition the values are: I-D Subscale, $r=.70$, $.42$; P-A Subscale, $r=.60$, $.36$; N-A Subscale, $r=.60$, $.41$; and Ma Subscale, $r=.60$, $.43$. Table K in Appendix B presents the values including systematic rater bias in the error.

Construct Validation of the CACL

One construct validity estimate of the CACL was obtained from intake ratings on the CACL and the results of the MMPI, when both were administered concurrently. The descriptive statistics for the CACL and MMPI are presented in Table 2, while the intercorrelations are presented in Table 3.

TABLE 2
DESCRIPTIVE STATISTICS FOR CONCURRENT VALIDITY STUDY

Variable	Mean	Standard Deviation	Number
CACL PA	46.89	4.95	140
CACL ID	49.59	6.32	140
CACL NA	46.46	5.34	140
CACL Ma	46.79	4.10	140
MMPI 1	5.79	4.58	140
MMPI F	15.35	10.76	140
MMPI K	13.50	6.16	140
MMPI Hs	17.04	6.85	140
MMPI D	29.82	6.84	140
MMPI Hy	23.90	7.09	140
MMPI Pd	29.77	4.53	140
MMPI Mf	26.49	4.68	140
MMPI Pa	16.23	6.52	140
MMPI Pt	32.17	8.34	140
MMpi Sc	39.60	11.82	140
MMpi Ma	24.52	6.65	140
MMPI Si	31.16	10.63	140

TABLE 3

INTERCORRELATION MATRIX FOR CONCURRENT VALIDATION STUDY
(N=104)

Variable	Variable CACL					Variable MMPI										
	PA	ID	NA	PA	L	F	K	HS	D	Hy	Pd	Mc	Pa	Pt	Sc	Si
CACL PA	1.0															
CACL ID	.01	1.0														
CACL NA	.37	.48	1.0													
CACL Ma	.68	-.06	.36	1.0												
MMPI L	.11	-.04	-.03	.04	1.0											
MMPI F	.25	.38	.33	.18	-.13	1.0										
MMPI K	-.03	-.04	-.01	.02	.60	-.42	1.0									
MMPI HS	.09	.36	.36	.13	.10	.40	.07	1.0								
MMPI D	-.13	.37	.22	-.09	.08	.37	-.13	.59	1.0							
MMPI Hy	.03	.15	.29	.09	.36	.26	.22	.79	.64	1.0						
MMPI Pd	.16	.09	.24	.10	.16	.31	.07	.47	.48	.60	1.0					
MMPI Mc	.02	.09	.17	-.05	.16	.37	-.07	.27	.39	.36	.32	1.0				
MMPI Pa	.10	.25	.28	.14	-.01	.66	-.33	.45	.44	.37	.35	.42	1.0			
MMPI Pt	.02	.26	.30	.03	-.23	.71	-.27	.51	.61	.41	.39	.39	.52	1.0		
MMPI Sc	.13	.35	.40	.09	-.11	.86	-.30	.54	.54	.43	.42	.50	.73	.82	1.0	
MMPI Ma	.30	.12	.20	.17	.26	.44	-.08	.13	.01	.14	.20	.21	.45	.21	.43	1.0
MMPI Si	-.26	.20	.02	-.30	-.16	.36	-.40	.21	.54	.17	.27	.39	.36	.45	.48	.05 1.0

The results of the canonical correlation analysis which was performed on these data are presented in Table 4. The two canonical variates which appear in this table are the only two which produced a canonical correlation coefficient which was significant at least at the .05 level. The χ^2 values which are reported in this table test the significance of the cumulative canonical correlation coefficient as each canonical variate is removed. That is, the first χ^2 value tests the significance of all canonical correlation coefficients and the second χ^2 value tests the significance of all canonical correlation coefficients, after the first has been removed.

Table 5 includes the canonical weights for the MMPI and CACL subtest or both canonical variates. Table 6 includes the product-moment correlations between the subtests of those two instruments and each canonical variate.

As was mentioned in the methodology section, a redundancy analysis was performed to provide estimates of the variance shared between the MMPI and CACL. Two redundancy coefficients were calculated: one estimating the redundancy of the MMPI, given the CACL ($R_{\text{MMPI/CACL}}$), and the other, that of the CACL given the MMPI ($R_{\text{CACL/MMPI}}$). As Stewart and Love (1968) pointed out, in a case such as this, both of these estimates are necessary since the total variances of the two instruments are not equal. For this study, the $R_{\text{MMPI/CACL}}$ was equal to .095, and the $R_{\text{CACL/MMPI}}$ was equal to .199.

TABLE 4
RESULTS OF CANONICAL CORRELATION ANALYSIS OF
THE CACL AND MMPI

Canonical Variate	Eigenvalue	Canonical Correlation	Wilkes Lambda	χ^2	D.F.
1	.38	.61	.39	122.78	52*
2	.22	.47	.63	68.96	36*

* $p < .05$.

TABLE 5
 CANONICAL WEIGHTS OF MMPI AND CACL SUBTESTS FOR
 CANONICAL VARIATES 1 AND 2
 (N=140)

Subtest	Canonical Variate 1 - Weights	Canonical Variate 2 - Weights
CACL-PA	.144	1.028
IDEP	.962	- .134
NA	-.041	- .349
NA	.234	.012
MMPI - L	-.131	.433
F	.936	.504
K	.324	- .529
Hs	.651	.091
D	.537	- .201
Hy	-.534	- .196
Dd	-.001	.572
MF	-.022	.019
Pa	-.086	- .601
Pt	-.654	- .074
Sc	-.022	.008
Ma	.022	.138
Si	-.109	- .733

Criterion-Related Validity of the CACL

Several separate analyses were carried out to estimate the predictive validity of the CACL. First, a series of multiple regression analyses was carried out with the subtest scores on the CACL as the independent variable, and each measure of institutional disruptiveness as the dependent variable. In each analysis, the CACL subtest which theoretically should have shown the highest degree of association with the dependent variable was entered first in the regression. Because there was no rationale for the ordering of the remaining subtests, they were entered as a set on the second step.

Table 7 includes the descriptive statistics for this analysis, and Table 8 gives the intercorrelation matrix for all variables. The results of the multiple regression analysis are presented in Tables 9 through 12, inclusive. These tables include the results for suicide attempts, assaults, threats, and interactions, respectively.

The multiple regression analysis reported in Tables 9 through 12 was performed by entering first the CACL subscale which logically was considered the best predictor of each type of disruptive behavior. The other three subscales were entered as a set on a second step. Accordingly, these tables include the multiple R , R^2 , and F value testing the significance of the R^2 on steps one and two. In addition, the standardized partial regression coefficients

TABLE 6
 PRODUCT MOMENT CORRELATIONS BETWEEN SUBTESTS OF THE CACL
 AND MMPI AND CANONICAL VARIATES
 (N=139)

	<u>Canonical Variate 1</u>		<u>Canonical Variate 2</u>	
	CACL	MMPI	CACL	MMPI
CACL PA	.31	.19	.91	.42
CACL IDEP	.94	.58	-.27	-.12
CACL NA	.56	.34	-.05	-.02
CACL NA	.30	.19	.55	.25
MMPI L	-.01	-.03	.13	.28
MMPI F	.49	.81	.04	.07
MMPI K	-.09	-.15	.01	.03
MMPI Hs	.40	.65	-.07	-.16
MMPI D	.31	.50	-.19	-.42
MMPI Hy	.19	.31	.04	-.11
MMPI Pd	.14	.23	.08	.17
MMPI Mf	.13	.22	-.04	-.11
MMPI Pa	.32	.52	-.12	-.26
MMPI Pt	.27	.44	-.11	-.25
MMPI Sc	.43	.71	-.07	-.16
MMPI Ma	.21	.34	.17	.37
MMPI Si	.18	.29	-.28	-.62

TABLE 7
DESCRIPTIVE STATISTICS FOR PREDICTIVE VALIDITY STUDY

	Mean	Standard Deviation	<u>N</u>
Suicide Attempts	.12	1.14	104
Assaults	.60	1.15	104
Threats of Assault	1.08	1.67	104
Infractions of Rules	.34	.97	104
CACL PA	47.12	5.18	104
CACL IDEP	50.54	6.33	104
CACL NA	46.77	5.32	104
CACL Ma	47.03	4.38	104
MMPI L	5.76	5.13	104
MMPI F	16.58	10.96	104
MMPI K	13.22	6.05	104
MMPI Hs	17.54	7.15	104
MMPI D	25.64	6.72	104
MMPI Hy	24.26	7.56	104
MMPI Pd	29.89	4.62	104
MMPI Pf	26.89	4.67	104
MMPI Pa	16.59	6.54	104
MMPI Pt	32.53	8.46	104
MMPI Sc	40.64	12.34	104
MMPI Ma	24.72	7.23	104
MMPI Si	31.99	10.50	104

TABLE 8
INTERCORRELATION MATRIX FOR PREDICTIVE VALIDITY STUDY

Variable	Suicide Attempts	Threats of Assaults	Assaults	Other Infractions
CACL PA	.14	.29	.16	.13
CACL ID	-.05	.06	-.04	-.06
CACL NA	.14	.13	.03	.11
CACL Ma	-.08	.14	.03	.14
MMPI L	-.04	.12	.11	-.09
MMPI F	.16	.02	.05	.13
MMPI K	-.18	.13	.07	.06
MMPI Hs	.08	-.03	-.08	.11
MMPI D	.07	.18	-.09	.14
MMPI Hy	.07	.01	-.06	-.09
MMPI Pd	.16	.06	.05	.01
MMPI Mt	.12	.07	.01	.03
MMPI Pa	.23	.02	.01	-.14
MMPI Pt	.10	.05	.01	-.14
MMPI Sc	.18	.07	.05	-.08
MMPI Ma	.06	.24	.11	-.09
MMPI Si	.07	-.16	-.07	.07
Suicide Attempts	1.00	-	-	-
Threats of Assaults	.08	1.00	-	-
Assaults	.13	.40	1.00	-
Other Infractions	.21	.25	.27	1.00

TABLE 9
RESULTS OF MULTIPLE REGRESSION OF FREQUENCY
OF SUICIDE ATTEMPTS ON CACL SUBSCALES

CACL Subscale	Step	<u>R</u>	<u>R</u> ²	<u>F</u>	Beta	<u>F</u> (unique)
NA	1	.14	.02	2.09	.28	5.13*
MA	2				-.41	9.60
PA	2				.32	5.83*
IDEP	2	.35	.13	3.63	-.21	3.55

* $p < .05$, 1 and 99 df

TABLE 10
RESULTS OF MULTIPLE REGRESSION OF FREQUENCY
OF ASSAULTS ON CACL SUBSCALES

CACL Subscale	Step	<u>R</u>	<u>R</u> ²		Beta	<u>F</u> (unique)
PA	1	.16	.03	2.62	.25	3.40
MA	2				-.15	1.18
IDEP	2				-.07	.30
NA	2	.20	.04	.99	.02	.40

TABLE 11
RESULTS OF MULTIPLE REGRESSION OF FREQUENCY
OF THREATS OF ASSAULT ON CACL SUBSCALES

CACL Subscale	Step	<u>R</u>	<u>R</u> ²	<u>F</u>	Beta	<u>F</u> (unique)
FA	1	.29	.08	9.25**	.34	6.53*
MA	2					.92
IDEF	2					.96
NA	2	.31	.10	2.69	.10	.68

*P<.05, 1 and 99 df

TABLE 12
RESULTS OF MULTIPLE REGRESSION OF FREQUENCY
OF INFRACTIONS ON CACL SUBSCALES

CACL Subscale	Step	<u>R</u>	<u>R</u> ²	<u>F</u>	Beta	<u>F</u> (unique)
PA	1	.13	.02	1.85	.05	.13
MA	2				.05	.14
NA	2				.13	1.01
IDEF	2	.19	.04	.90	-.12	1.09

along with their corresponding tests of significance are reported for a model including all four subscales.

Relationship of the CACL to Crime

In order to assess the degree of relationship between the CACL and the presence of violence in the crime for which the subjects had been most recently convicted, a discriminant function analysis was performed. The CACL's subtests were the independent variables and charges at the time of arrest categorized into violent and nonviolent types constituted the grouping variable. Charges of property crime, drug charges, and others which did not involve physical contact were categorized as nonviolent. Those which involved a physical attack, battery, rape or murder were categorized as violent.

These data were analyzed by means of a discriminant function analysis, which provides weights for the independent variables such that the ratio of the sums of squares between groups to sums of squares within groups is maximized. In the procedure the dependent variable is categorical in nature, while the independent variables are continuous. This can be seen as a special case of the multiple regression procedure described previously, and therefore a multiple regression analysis was performed. The results of the analysis are presented in Table 13.

TABLE 13
RESULTS FOR DISCRIMINANT FUNCTION ANALYSIS OF
CRIME TYPE
(N=104)

Source	<u>D.F.</u>	<u>S.S.</u>	<u>M.S.</u>	<u>F</u>
Regression	4	1.16	.29	1.18(n.s.)
Residual	99	24.22	.24	

Summary

In summary, this chapter included a presentation of the results of two inter-rater reliability studies on the CACL, as well as a number of validation procedures. These included assessments of the CACL's predictive validity in terms of four types of institutional disruptiveness, a concurrent validation with the MMPI, as well as a study of the instruments "post-dictive" relationship to the degree of violence involved in the subjects most recent crime. The following chapter will relate those findings to the overall utility of the CACL as a classification instrument in correctional settings.

CHAPTER V

DISCUSSION

The purpose of this study has been to investigate the psychometric properties of the Quay Correctional Adjustment Checklist (CACL), which is a empirically derived classification instrument used with incarcerated criminals. In order to provide estimates of the instrument's reliability and validity, data were gathered on a sample of males who were being treated in a maximum security mental hospital.

Because an average of three raters' scores was used for placement decisions within the institution, intraclass correlation coefficients were calculated for this average. For the purpose of comparison, these coefficients were also calculated for single raters. Since the CACL may be used for either absolute or comparative decisions, reliability coefficients were also calculated which included systematic rater bias in the error term (when the CACL is used for absolute decisions) as well as deleting it (when the CACL is used for comparative decisions). These coefficients were calculated because although this study does not address the various decision rules or cutting scores which may be used in classifying individuals with the CACL, some readers may wish to use the CACL to compare individuals

rather than making absolute placement decisions. In any case, reliability estimates will need to be calculated for the CACL for decision rules different than that used here.

Two estimates of the CACL's reliability were obtained, one based on ratings of the subjects during the "intake and diagnostic" phase of treatment, and one after a longer period of observation and more controlled training of the raters. Both of these studies provided reliability estimates for the average of three raters which were larger than .50 in every instance except one.

In order to assess the usefulness of the CACL with forensic psychiatric patients, three validity studies were conducted. First, a concurrent validation study, relating scores on the MMPI to those on the CACL was conducted. Second, subscale scores on the CACL were used to predict several types of disruptive behavior within the institution. Finally, scores on the CACL were related to the presence of violence in the subjects.

This section contains a synthesis and interpretation, as well as a summary of results. Particularly in the validation analyses, an effort is made to place the results in a framework of meaning and applicability in other settings. The final section of this chapter contains suggestions for future research to improve the utility of those results obtained in this study.

Construct Validation of the CACL

The first inter-rater reliability study was based on data collected during the intake and evaluation phase of treatment. The subjects varied more in their crime types than did those in the second study and included a large percentage of individuals diagnosed as psychotic. The sex offenders included in the second study were by definition, nonpsychotic.

Since the sample used in the second study was more homogeneous than that in the first, the reliability estimates obtained would tend to be somewhat lower than those obtained at intake. Thus, the reliability estimates from the "controlled" study do not represent the maximum possible for the instrument. Despite the homogeneity of the sample, three of the four scales on the CACL showed increases in the magnitude of obtained reliability estimates when the raters were thoroughly trained¹ in the definitions of terms and when a longer observation period was available before rating. Those subscales showing increases in the magnitude of obtained reliability estimates from "intake" to "controlled" settings, showed small gains in reliability. Estimates for single raters were much lower than those based on the average of three raters.

Construct Validation of the CACL

The canonical correlation analysis relating the CACL and MMPI was intended to determine the overall relationship

between the two instruments, based on the analysis of data from a sample of 140 individuals who were assessed with both instruments during the first week of their stay at the institution. Two canonical variates were derived, and a product moment correlation coefficient between the weighted combination of CACL and MMPI subscale scores was calculated for each.

The canonical correlation coefficients of .61 and .47 do not represent estimates of relationship between the total variance of each measure, but rather involve only that variance in each which was included in the particular linear combination of subscales (Stewart & Love, 1968). Thus, these coefficients cannot be squared to determine the percentage of total variance common to both measures.

The canonical correlation coefficients for both canonical variates are significant at the .05 level when tested with a chi-square statistic². Thus, for the two derived variates, a statistically significant relationship exists between the subscales of the MMPI and CACL.

In addition to the canonical variate analysis, a redundancy analysis was performed in order to estimate the degree of congruence or overlap between the two measures. Two redundancy coefficients were derived by this analysis, one estimating the redundancy in the CACL given the MMPI, and a second estimating the redundancy in the MMPI, given the CACL. The values for these coefficients are .19 and

.12, respectively. The magnitude of these coefficients indicates a very modest relationship between the two sets of variables. In order to assess the concurrent validity of the CACL, it is necessary to examine the unique contribution of each subscale to the canonical variate in question, as well as to interpret the variates through the examination of the correlations between the original variables and their canonical weights.

High positive weightings on the two canonical variates which were derived in this analysis appear to describe two distinct types of individuals within the population of the North Florida Evaluation and Treatment Center. That is, ratings on the CACL subscales do not correspond to four distinct patterns of canonical weights on the MMPI. Instead, the two variates which emerge load heavily on the Psychopathic-Aggressive (PA) and Manipulative (Ma) subscales in one case, and on the Immature-Dependent (ID) and Neurotic-Anxious (NA) subscales in the second. For the purpose of convenience these canonical variates will be labelled according to the CACL subtest on which they load most heavily. The first variate will be called "Immature-Dependent" and the second, "Psychopathic-Aggressive."

The "Immature-Dependent" variate correlates highly (.60 and above) with MMPI subtests which relate to unusual responses (subscale F), bodily discomfort or illness (subscale Hs), and bizarre or psychotic symptoms (subscale Sc).

This variate also correlated .40 to .50 with subscales assessing hostility and suspiciousness (P-A) and overt symptoms of depression (subscale D).

The "Psychopathic-Aggressive" variate correlated +.55 with the Ma subscale. The MMPI subscales which showed correlation of the largest magnitude were those measuring social introversion (Si subscale) where $r = .61$; and the depression (D) subscale, where $r = .42$.

Although dealing with a different population, Brown (1968) described the characteristics of a group of subjects at the Robert F. Kennedy Youth Center, who had been classified with the CACL as "inadequate-immature" delinquents. These descriptions correspond closely to the content of the "Immature-Dependent" variate.

Brown (1968) also noted that such an individual is described as ". . . lazy, immature, a daydreamer, reticent, showing a lack of interest in things. His relationships are characterized by resentment (towards authority figures) or dependency. . ." (p. 3).

Similarly, a group of individuals who are labelled as "psychopathic-aggressive" in the same study are described as "assaultive, cruel, defiant . . . wiley, deceitful and very untrustworthy . . . (such individuals) discount past mistakes and see their future without problems and themselves as great successes. . ." (p. 7). This description matches that given by Dahlstrom in the MMPI Handbook

(1972) for individuals with low Si scale scores. He said that such individuals tend to be "active and vigorous, and competitive with their peers. They are persuasive and often win others over to their viewpoint. They also manipulated others in attempting to gain their own ends . . . they appeared unable to delay gratification and often acted with insufficient thought or deliberation . . . (which) . . . led to a destructive aggressiveness or hostility in their personal relations (p. 172).

It seems likely that this variate describes a group of individuals who tend to deny depression, are active and aggressive, and who have low impulse control. They tend to control others through verbal behavior or physical violence, and to have a great deal of energy and a rapid flight of ideas.

This description fits those persons who were admitted to NFETC because they were too assaultive or dangerous to others to be kept in other institutions. Their tendency to act out under minor stress and their need to control others were the likely cause of their confinement at the institution.

The immature-dependent variate may well describe those individuals who are overtly psychotic but not highly agitated or assaultive. It also may include the individuals who were committed for evaluation and who are "faking bad" by pretending to be psychotic and or physically ill. In any case,

such persons are seen as lethargic, withdrawn and passive individuals who depend highly on others to meet their needs.

Whatever the reason, with this population, the CACL did not aggregate the subjects into four distinct groups. Rather, two groups emerged on both the CACL and MMPI, which seemed to differ primarily on the dimensions of activity control of others, and somatic complaints. For the "psychopathic aggressive" variate, this corresponds to Quay's earlier finding of the high degree of relationship between the PA and Ma subscales (Quay, 1971, p. 7).

Criterion Validity of the CACL

The second set of validity analyses on the CACL are concerned with the relationship between scores on its subtests and measures of disruptive behavior within the institution. Although the primary concern in this case is with predictive validity, inferences about construct validity may also be made since the dependent variables are of interest in regard to their logical relationship to the CACL subtests, as well as being of concern in themselves.

For example, we would expect the Psychopathic-Aggressive subscale to show a positive relationship of greater magnitude to threats of physical assault than do the other subscales, and this is in fact the case. This type of relationship provides partial confirmation of the validity of the trait names underlying the subtests of the

CACL, and follows the distinction between predictive and construct validation made in the APA Standards for Psychological Tests.

It should be noted that the four criterion measures which were chosen for the study showed very little variability. This may be because every effort was made to prevent the occurrence of those behaviors, and because many of the subjects were on their best behavior during the first sixty days of confinement. Whatever the reason, the lack of variability in the criterion tends to "cause" a decrease in validity estimates such as those presented here.

Each category of disruptive behavior is described separately, and the results of the overall multiple regression of the CACL on each will be presented. Additionally, the multiple regression analysis is discussed in terms of the contribution of each individual subtest to the variance in each dependent variable.

Suicide Attempts

The results of the multiple regression analysis indicate that the amount of variance in suicide attempts which is predicted by the CACL as a total test is significant at the .05 level. In addition, the unique contribution of each subtest is also significant at that level. It should be noted that with all the subtests in the prediction equation, only 13% of the variance in suicide attempts was predicted by the CACL.

The Neurotic-Anxious subscale showed the largest degree of association with suicide attempts, and this is logically consistent with the hypothetical trait measured by this subscale. Individuals of this type have been characterized by Brown (1968) as ". . . fearful, anxious, withdrawn, hypersensitive, self-conscious, having feelings of inferiority and lacking self-confidence. . ." (p. 5).

The question can be raised at this point as to the intent underlying the suicide attempts in question. As Samenow (1978) has pointed out, mental hospitals are preferable to prisons in terms of creature comforts. In his study, many individuals feigned psychopathology to be transferred to a more comfortable environment (usually a hospital). Suicide threats and gestures were often used by the inmates to prevent their return to prison.

Thus, the suicide attempts may have been either sincere or an effort to maintain status as a "patient" in need of treatment. Although it is not possible to retrospectively determine the reasons for such behavior, the results of the multiple regression analysis provide some confirmation of this hypothesis. The two highest positive beta weights in the prediction equation are for the Psychopathic-Aggressive and Neurotic-Anxious subscales. Thus, it may be that although these subscales relate significantly to the behavior in question, they may be discriminating between the two types of behavior (i.e., manipulative versus self-destructive).

The relatively high negative weighting on the Manipulative subscale ($B = -.41$) seems to be inconsistent with this hypothesis. However, the items on this subscale reflect behaviors such as lying and cheating, rather than less obvious manipulations. It is possible that such individuals have a repertoire of manipulative behaviors which are more effective than fake suicide attempts.

Threats of Assault

The overall F ratio for the multiple regression of the CACL on threats of assault is statistically significant at the .05 level. On inspection, however, only the Psychopathic-Aggressive subscale appears to accounting for a significant unique amount of variance to the dependent variable.

The PA subscale predicts eight percent of the variance in threats of assault. The high loading on the PA subscale provides further evidence for the nature of the theoretical trait which it purports to measure. That is, we would expect such individuals to be aggressive, hostile and domineering in social interactions. Because such persons are thought to be easily frustrated, they are presumed to react to minor stress with a variety of aggressive manipulations, including threats of physical violence such as those recorded by the staff.

It seems likely that threats of physical assault are less determined by the characteristics of the residents at the time of evaluation than they are determined by aspects

of the environment of the time. Assault or threats of assault are intrinsically social behaviors, while suicide attempts are most often carried out in private. Thus, threats of assault are also partially determined by the behavior of the person being threatened.

Assaults

The overall F ratio for the multiple regression analysis of the CACL on instances of physical attack or assault is not statistically significant at the .05 level. This may well be due to the dyadic nature of the social interaction being measured. Although threats of assault may be seen as a manipulative style, a physical act of violence was usually followed by close confinement in a seclusion room. Again, repeated acts of violence were considered as grounds for transfer from the hospital, and may not have occurred with any frequency during the first two weeks of confinement. All of these factors may have contributed to the inability of the CACL to predict such behaviors. [It is of interest to note at this point that a recent review of dangerousness (Gottfredson, 1971) in a variety of settings showed a plethora of negative results in studies of individual characteristics, perhaps for the same reasons.]

Infractions of Rules

The overall multiple regression analysis for the CACL on minor infractions of rules was not significant at the .05

level. None of the CACL subtests appears to relate significantly to the frequency of minor acts which were contrary to rules of the institution or unit where the residents were housed.

This may be the result of several factors. First, the frequency of these infractions was the lowest of any of the recorded disruptive behaviors, having a mean recorded incidence of .34 for the sixty-day period. These infractions also showed the smallest variance of any of the recorded disruptive behaviors. This lack of variance may account for the minimal prediction of the CACL. Also, such behaviors may be the result of ignorance of the rules rather than of a prior condition, as measured by the CACL.

Relationship of the CACL to Crime Type

A final validity estimate on the CACL was derived from a discriminant function analysis in which the instrument's subscale scores were related to the presence of violence in the most recent crime for which they had been convicted. The results of this analysis were not statistically significant at the .05 level.

Several factors may be responsible for this apparent lack of validity. First, the CACL was unable to predict violence within the institution, perhaps because of the short-time period that was involved in this study. As was mentioned before, violence is an interaction between two

persons, and cannot be predicted well based on a knowledge of the characteristics of one individual. Second, the presence of violence was assessed by categorizing the crimes for which each subject had been convicted. Since these convictions often were based on plea bargaining, they may not have accurately measured the amount of violence present when the crime was actually committed.

Summary of Psychometric Evaluation

The purpose of this study was to evaluate the CACL in terms of its psychometric properties as measured in a variety of ways. Based on the various analyses which were performed, as well as the general characteristics of the instrument, several conclusions can be reached in regard to this evaluation.

First, the instrument is polythetic in nature and provides rankings of individuals along several "behavioral dimensions." Although these dimensions were originally derived from a factor analysis, they have never been conclusively shown to be independent in later studies. This study and others have shown two clusters of traits rather than the four which were originally derived. The fact that these two groups of behaviors have been found in three separate studies with distinctly different populations indicates that they may well reflect actual methods of coping with a prison or hospital environment.

Second, the CACL can be used to provide scores which produce inter-rater reliability estimates in the range of .60 to .70, for the average of three raters. These estimates are much lower for a single rater. It does not appear to be necessary for the raters to observe the subjects for two weeks to provide reliable scores on the instrument, but the short observation period used in these studies (3-6 days) may well have limited the validity estimates which were obtained. That is, this time period may not have been long enough to observe characteristic behavior patterns, since many persons may have been on their best behavior at the time of admission.

Third, when used with a sample of mentally or behaviorally disordered individuals the CACL appears to be measuring some of the same traits as the MMPI. A construct validation study using a canonical correlation showed two underlying clusters of traits in this sample, and provided canonical correlation values of .61 and .47 for the two groups of subtests. For this sample and others, the CACL seems to be measuring dominance, aggression and mania in one dimension and feelings of distress, depression, social withdrawal and anxiety in the second.

Since the Psychopathic-Aggressive and Manipulative subscales both load highly on the first dimension and the Immature-Dependent and Neurotic-Anxious subscales load on the second, it is possible that the raters were responding

to more gross behavioral evidence than is desirable, and were tending to rate the subjects globally rather than specifically.

Fourth, in the sample described above, the CACL shows a statistically significant relationship to the frequency of suicide attempts and to threats of violence, but not to other measures of disruptive behavior in the institution. The subtests which have the highest unique contribution to the prediction of those behaviors are the ones which would be expected to do so. That is, suicide attempts are most highly related to high scores on the Neurotic-Anxious and Psychopathic-Aggressive subscales, perhaps corresponding to real and feigned suicidality. The Psychopathic-Aggressive subscale showed the largest relationship in threats of violence, as would be expected.

Finally, the CACL showed no statistically significant relationship to be presence of violence in the subjects most recent crime. However, the Manipulative subtest had a positive correlation of .17 with the presence of violence. While this is not statistically significant, it does provide some basis for speculation. It appears that more manipulative individuals tend to have more violence in their crime types than do other individuals.

In general, the CACL appears to have some potential for useful classification in maximum security mental hospitals. Its inter-rater reliability is so low that it should not be

used by a single rater to arrive at placement decisions. When three raters are used, the reliability estimates increase somewhat, but still do not provide much basis for placement decisions in the absence of other information. Although its value is limited by the extent of the raters' observation, it has the advantage of not being biased by the same response sets which influence self-report inventories such as the MMPI.

Recommendations

Several changes in the CACL might improve its reliability and general utility. It would be helpful if more precise definitions of the terms used were provided. During the course of training, several raters complained that no standards were given for decisions as to whether a particular behavior was included within a category on the CACL.

Also, the instrument could be converted to a rating scale rather than a checklist. This would allow for more precise description of each individual than is currently possible, and by increasing the inter-individual variance would allow for more meaningful discriminations between individuals.

Since the results of this study indicate that the CACL appears to have both construct and predictive validity, further efforts should be made to measure the stability of the behavioral traits which it measures. In this way, the instrument could be related to treatment outcomes and used

for measuring individual change across time. Since it is a nonreactive measure, it has the potential for serial administration without the reactive effects of other classification instruments.

The question of the actual number of traits being measured by the CACL needs to be answered. It is possible that more precise behavioral definitions would allow for the assessment of whether two or four traits are being assessed. A larger sample of individuals should be assessed by raters who have been well trained, and who have observed the subjects in a variety of settings. A factor analysis of these results would provide more definitive evidence of this question.

In general the CACL meets many of the requirements for an effective classification instrument. It provides a method for assessing behavioral styles in incarceration that have implications for both management and treatment. The lack of inter-rater reliability probably sets a limit on the validity of the instrument, and until this problem is improved it should be used with great caution.

APPENDIX A

CORRECTIONAL ADJUSTMENT CHECKLIST

APPENDIX A

CORRECTIONAL ADJUSTMENT CHECKLIST

Marked for Final Factor Scales

- Scale I (Aggressive--Psychopathic) (N=18)
- Scale II (Immature--Dependent) (N=11)
- + Scale III (Neurotic--Anxious) (N=7)
- Scale IV (Manipulative) (N=5)

	Col	No.	
III +	(17)	0 1	1. Worried, anxious
II -	(18)	0 1	2. Tries, but cannot seem to follow directions
III +	(19)	0 1	3. Tense, unable to relax
II -	(21)	0 1	4. Socially withdrawn
III +	(22)	0 1	5. Continually asks for help from staff
I -	(24)	0 1	6. Gets along with the hoods
II -	(25)	0 1	7. Seems to take no pleasure in anything
III +	(26)	0 1	8. Jittery, jumpy; seems afraid
I -	(27)	0 1	9. Uses leisure time to cause trouble
I -	(28)	0 1	10. Continually uses profane language; curses and swears
III +	(29)	0 1	11. Easily upset
II -	(30)	0 1	12. Sluggish and drowsy
I -	(31)	0 1	13. Cannot be trusted at all
II -	(32)	0 1	14. Moody; brooding
I -	(34)	0 1	15. Needs constant supervision
I -	(35)	0 1	16. Victimizes weaker inmates
II -	(36)	0 1	17. Seems dull and unintelligent
I -	(38)	0 1	18. Is an agitator about race
IV -	(40)	0 1	19. Continually tries to con staff
I -	(41)	0 1	20. Impulsive; unpredictable
III +	(42)	0 1	21. Afraid of other inmates

	<u>Col. No.</u>			
I -	(43)	0	1	22. Seems to seek excitement
II -	(44)	0	1	23. Never seems happy
IV -	(46)	0	1	24. Doesn't trust staff
II -	(49)	0	1	25. Passive; easily led
I -	(53)	0	1	26. Talks aggressively to other inmates
I -	(55)	0	1	27. Accepts no blame for any of his troubles
IV -	(56)	0	1	28. Continually complains; accuses staff of unfairness
II -	(59)	0	1	29. Daydreams; seems to be mentally off in space
I -	(62)	0	1	30. Talks aggressively to staff
I -	(64)	0	1	31. Has a quick temper
I -	(65)	0	1	32. Obviously holds grudges; seeks to "get even"
II -	(68)	0	1	33. Inattentive; seems preoccupied
IV -	(70)	0	1	34. Attempts to play staff against one another
II -	(71)	0	1	35. Passively resistant; has to be forced to participate
I -	(72)	0	1	36. Tries to form a clique
I -	(73)	0	1	37. Openly defies regulations and rules
III +	(74)	0	1	38. Often sad and depressed
I -	(75)	0	1	39. Stirs up trouble among inmates
I -	(76)	0	1	40. Aiding or abetting others in breaking the rules
IV -	(77)	0	1	41. Considers himself unjustly confined

•

APPENDIX B

SUMMARY TABLES FOR INTER-RATER
RELIABILITY STUDIES

•

TABLE 14
DESCRIPTIVE STATISTICS FOR INTER-RATER RELIABILITY
STUDY: "INTAKE" CONDITION

Variable	Mean	Standard Deviation	Number
CACL PA	46.89	4.95	140
CACL ID	49.59	6.32	140
CACL NA	46.46	5.34	140
CACL Ma	46.79	4.10	140

TABLE 15
DESCRIPTIVE STATISTICS FOR INTER-RATER RELIABILITY
STUDY: "CONTROLLED" CONDITION

Variable	Mean	Standard Deviation	Number
CACL PA	47.12	5.18	69
CACL ID	50.54	6.33	69
CACL NA	46.77	5.32	69
CACL Ma	47.03	4.38	69

TABLE 16
 ANALYSIS OF VARIANCE TABLE FOR CACL PA
 "CONTROLLED" CONDITION INTER-RATER
 RELIABILITY STUDY

Source	Sums of Squares	D.F.	Mean Square
Rater	173.07	2	86.54
Subjects	2483.94	22	112.91
Residual	1216.93	44	27.66
Total	3873.94	68	

TABLE 17
 ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL ID
 "CONTROLLED" CONDITION INTER-RATER
 RELIABILITY STUDY

Source	Sum of Squares	D.F.	Mean Square
Rater	314.46	2	157.23
Subjects	1837.28	22	83.51
Residual	2304.20	44	52.37
Total	4455.94	68	

TABLE 18
 ANALYSIS OF VARIANCE TABLE FOR CACL NA
 "CONTROLLED" CONDITION INTER-RATER
 RELIABILITY STUDY

Source	Sum of Squares	<u>D.F.</u>	Mean Square
Rater	71.04	2	35.52
Subjects	2279.30	22	103.61
Residual	1252.96	44	28.47
Total	3603.30	68	

TABLE 19
 ANALYSIS OF VARIANCE TABLE FOR CACL Ma
 "CONTROLLED" CONDITION INTER-RATER
 RELIABILITY STUDY

Source	Sum of Squares	<u>D.F.</u>	Mean Square
Rater	81.48	2	40.74
Subjects	1954.44	22	88.84
Residual	880.52	44	20.01
Total	2916.44	68	

TABLE 20
ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL PA
"INTAKE" CONDITION INTER-RATER
RELIABILITY STUDY

Source	Sum of Squares	<u>D.F.</u>	Mean Square
Raters	32.43	2	16.21
Subjects	8298.87	103	80.57
Residual	6703.57	206	32.54
Total	15034.87	311	

TABLE 21
ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL ID
"INTAKE" CONDITION INTER-RATER
RELIABILITY STUDY

Source	Sum of Squares	<u>D.F.</u>	Mean Square
Raters	92.55	2	46.28
Subjects	12372.21	103	120.12
Residual	7392.78	206	35.89
Total	19857.54	311	

TABLE 22
ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL NA
"INTAKE" CONDITION INTER-RATER
RELIABILITY STUDY

Source	Sum of Squares	<u>D.F.</u>	Mean Square
Rater	7.71	2	3.86
Subjects	8733.51	103	84.79
Residual	6907.62	206	33.54
Total	15648.84	311	

TABLE 23
ANALYSIS OF VARIANCE SUMMARY TABLE FOR CACL Ma
"INTAKE" CONDITION INTER-RATER
RELIABILITY STUDY

Source	Sum of Squares	<u>D.F.</u>	Mean Square
Raters	13.97	2	6.98
Subjects	5933.28	103	57.60
Residual	4815.36	206	23.38
Total	10762.61	311	

TABLE 24
 INTER-RATER RELIABILITY COEFFICIENTS FOR INTAKE
 AND CONTROLLED CONDITIONS, INCLUDING SYSTEMATIC
 RATER BIAS IN THE ERROR TERM

	Intake		Controlled	
	Average of Three Raters	Single Rater	Average of Three Raters	Single Rater
CACL ID	.69	.40	.34	.26
CACL PA	.56	.33	.71	.49
CACL NA	.57	.39	.72	.44
CACL Ma	.59	.43	.71	.58

REFERENCES

- Abrahamson, D. The psychology of crime. New York: Holt, Rinehart and Winston, 1960.
- Alexander, F. G. Roots of crime: Psychoanalytic Studies. New York, N.Y.: Alfred A. Knopf, 1935.
- Alexander, F., & Staub, H. The criminal, the judge and the public. Glencoe, Ill.: Glencoe Publishing Co., 1956.
- American Psychological Association. Standards for educational and psychological tests and manuals. Washington, D.C.: American Psychological Association, 1974.
- Anastasi, A. Psychological testing (4th ed.). New York: Macmillan Publishing Co., 1976.
- Argyle, D. C. A new approach to the classification of delinquents with implications for treatment. California State Board of Corrections. Monograph 2, 1961, 15-26.
- Bartko, J. J. On various intraclass correlation reliability coefficients. Psychological Bulletin, 1966, 83, 762-765.
- Blackburn, F. An empirical classification of psychopathic personality. British Journal of Psychiatry, 1975, 127, 456-460.
- Brogden, H. E., & Taylor, E. K. The theory and classification of criterion bias. Educational and Psychological Measurement, 1950, 10, 159-186.
- Bromberg, W., & Thompson, C. B. The relationship of psychosis, mental defect and personality types to crime. Journal of Criminal Law and Criminology, 1937, 28, 70-89.
- Brown, D. E. The Robert Kennedy youth center: An interim report. Washington, D.C.: U.S. Department of Justice, 1968.
- Brown, F. G. Principles of educational and psychological testing. Hinsdale, Ill.: Dryden Press, 1970.
- Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multi-trait, multi-method technique. Psychological Bulletin, 1959, 56, 81-105.

- Cattell, R. B. Validity and reliability: A proposed more basic set of concepts. Journal of Educational Psychology, 1964, 55, 1-22.
- Clinnard, M. B. Sociology of deviant behavior. New York, N.Y.: Harper and Row Publishers, 1963.
- Clinnard, M. B., & Quinney, R. Criminal behavior systems: A typology. New York: Holt, Rinehart and Winston, 1973.
- Clinnard, M. B., & Quinney, R. Criminal behavior systems (2nd ed.). New York: Holt, Rinehart and Winston, 1967.
- Cronbach, L. J. Test "reliability": Its meaning and determination. Psychometrika, 1947, 12, 1-16.
- Cronbach, L. J. Coefficient alpha and the interval structure of tests. Psychometrika, 1951, 16, 297-334.
- Cronbach, L. J. Essentials of psychological testing (2nd ed.). New York: Harper and Row, 1960.
- Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.
- Cureton, E. E. Validity, reliability, and baloney. Educational and Psychological Measurement, 1950, 10 94-96.
- Cureton, E. E. The definition and estimation of test reliability. Educational and Psychological Measurement, 1958, 18, 715-738.
- Dahlstrom, W. G. An MMPI handbook (Vol. 1). Clinical interpretations. Minneapolis: University of Minnesota Press, 1972.
- Dahlstrom, W. G.; Welsh, G. S.; & Dahlstrom, L. F. An MMPI handbook (Vol. 2). Minneapolis, Minnesota: University of Minnesota Press, 1972.
- Driver, E. A critique of typologies in criminology. Sociological Quarterly, 1968, 9, 356-373.
- Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424.

- Ebel, R. L. Must all tests be valid? American Psychologist, 1961, 16, 640-647.
- Ferdinand, T. N. Typologies of delinquency. New York: Random House, 1966.
- Fisher, S. Varieties of juvenile delinquency. British Journal of Criminology, 1962, 2, 251-261.
- Frick, T., & Semmel, M. Observer agreement and reliabilities of classroom observational data. Review of Educational Research, 1978, 48(1), 157-184.
- Gaion, L. Criterion-related validity. Educational and Psychological Measurement, 1974, 32, 316-326.
- Gall, F. J. Craniology, and new discoveries about the head, the brain and the organs. Paris, France: publisher unknown, 1807.
- Gibbons, D. C. Changing the lawbreaker. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965.
- Gibbons, D. C. Society, crime and criminal careers (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.
- Gibbons, D. C. Offender typologies: Two decades later. British Journal of Criminology, 1975, 15(2), 211-221.
- Glaser, D. The new correctional era-implications for manpower and training. Crime and Delinquency, 1964, 12, 1-26.
- Glueck, S., & Glueck, E. Predicting delinquency and crime. Cambridge: Harvard University Press, 1959.
- Gottfredson, D. M. The base expectancy approach. In N. Johnson, L. Savirz, & M. Wolfgang (Eds.), The sociology of punishment and correction (2nd ed.). New York: John Wiley and Sons, 1971.
- Guikeson, H. Intrinsic validity. American Psychologist, 1960, 5, 511-517.
- Guilford, J. P. Factor analysis in a test development program. Psychological Review, 1948, 55, 479-494.
- Guze, S. B. Criminality and psychiatric disorders. New York: Oxford University Press, 1976.

- Haggard, E. A. Intraclass correlation and the analysis of variance. New York: Dryden Publishers, 1958.
- Hewitt, L. E., & Jenkins, R. L. Functional patterns of maladjustment. Springfield, Ill: State of Illinois, 1946.
- Hood, R., & Sparks, R. Key issues in criminology. New York: McGraw-Hill, 1970.
- Hoyt, C. J. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.
- Hunt, D., & Hardt, L. Developmental state, delinquency and differential treatment. Journal of Research in Crime and Delinquency, 1965, 23, 20-31.
- Hurwitz, L. Three delinquent types: A multivariate analysis. Journal of Criminal Law, Criminology and Police Science, 1965, 56, 328,334.
- Jenkins, R. A., & Hewitt, L. C. Types of personality structure encountered in child guidance clinics. American Journal of Orthopsychiatry, 1949, 14, 84-94.
- Jenkins, R. L., & Glickman, S. Patterns of personality organization among delinquents. Nervous Child, 1947, 6, 329-339.
- Jesness, C. The Preston typology study. British Journal of Crime and Criminology, 1959, 23, 112-128.
- Kerlinger, F. W., & Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston, 1973.
- Killinger, G., & Cromwell, P. (Eds.). Penology. St. Paul, Minn.: West Publishing Co., 1973.
- Kinch, J. W. Continuities in the study of criminal types. Journal of Criminal Law, Criminology, and Police Science, 1962, 53, 323-328.
- Kinch, J. W. Continuities in the study of delinquent types. Journal of Criminology Law, Criminology and Police Science, 1963, 54, 296-307.
- Kretschmer, E. Physique and character. London: W. J. H. Sprott, Publishers, 1925.

- Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. Psychometrika, 1937, 2, 151-160.
- Lambroso, C. Crime: Its causes and remedies. Boston, Mass.: Little, Brown and Co., 1911.
- Loveland, F. Classification in the prison system. In P. L. Tappan (Ed.), Contemporary corrections. New York: McGraw-Hill, 1951, 91.
- Mack, J. L. The MMPI and recidivism. Journal of Abnormal Psychology, 1969, 74, 612-614.
- Maddocks, P. D. A five-year follow-up of untreated psychopaths. British Journal of Psychiatry, 1970, 116, 511-515.
- Magnusson, D. Test theory, translated by Hunter Mahon. Reading, Mass.: Addison Wesley Co., 1967.
- McGaw, B. L.; Wardrap, J. L.; and Burda, M. A. Classroom observation schemes: Where are the errors? American Educational Research Journal, 1972, 9(1), 13-27.
- Meehl, P. E. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. Psychological Bulletin, 1955, 52, 194-216.
- Megargee, E. I. The need for a new classification system. Criminal Justice and Behavior, June 1977, 4(2), 107-113.
- Magargee, E. I., & Bohn, M. J. Empirically derived characteristics of the ten types. Criminal Justice and Behavior, June 1977, 4(2), 149-210.
- Megargee, E. I.; Meyer, J.; Darhut, B.; & Bohn, M. J. A new classification system for offenders. Criminal Justice and Behavior, 1977, 4(2), 107-214.
- Messick, S., & Jackson, D. N. Problems in human assessment. New York: McGraw-Hill, Inc., 1967.
- Monachesi, E. P., & Hathaway, S. R. The personality of delinquents. In MMPI: Research development and clinical applications. Minneapolis: University of Minnesota Press, 1972.

- Morris, A. The comprehensive classification of adult offenders. Journal of Criminal Law, Criminology and Police Science, 1965, 36, 197-202.
- Mosier, C. I. A critical examination of the concepts of face validity. Educational and Psychological Measurement, 1957, 7, 191-205.
- National Institute of Mental Health. Typological approaches and delinquency control: An interim report. Washington, D.C.: U.S. Government of Health, Education and Welfare, 1967.
- Palmer, J., & Carlson, P. Problems with the use of regression analysis in prediction studies. Journal of Research in Crime and Delinquency, 1966, 13(1), 64-79.
- Palmer, J. O. The psychological assessment of children. New York: Wiley Publishers, 1970.
- Panton, J. H. The identification of predispositional factors in self-mutilation within a state prison population. Journal of Clinical Psychology, 1965, 18, 63-67.
- Panton, J. H. The longitudinal effects of first incarceration on MMPI profiles. Unpublished paper. Raleigh, N.C.: North Carolina Department of Social Rehabilitation, 1966.
- Panton, J. H. The identification of predispositional factors influencing prison adjustment. Unpublished paper. Raleigh, N.C.: North Carolina Department of Social Rehabilitation, 1968.
- Panton, J. H. Manual for a prison classification inventory for the MMPI. Raleigh, N.C.: Department of Social Rehabilitation and Control. 1970.
- Peterson, H. R.; Quay, H. C.; & Tiffany, T. L. Personality factors related to juvenile delinquency. Child Development, 1961, 32, 355-372.
- Peterson, J.; Quay, J.; & Cameron, H. Personality and background factors in juvenile delinquency as inferred from questionnaire responses. Journal of Consulting Psychology, 1959, 23, 395-399.
- Peterson, R. A.; Pittman, D. J.; and O'Neal, P. Stabilities in deviance: A study of assaultive and non-assaultive offenders. Journal of Criminal Law, Criminal and Police Science, March 1962, 44-48.

- Quay, H. Personality dimensions in delinquency males inferred from the factor analysis of behavior ratings. Journal of Research in Crime and Delinquency, 1964, 24, 33-37.
- Quay, H. C. The differential behavioral classification of the adult male offender: Interim results and procedures. Unpublished report, presented to the National Institute of Mental Health, Washington, D.C., 1971.
- Quinney, R. The social reality of crime. Boston, Mass.: Little, Brown and Co., 1970.
- Quinney, R. Criminology: Analysis and critique of crime in America. Boston, Mass.: Little, Brown and Co., 1972.
- Reckless, W. C. The crime problem (4th ed.). New York, N.Y.: Appleton-Century-Crofts, 1967.
- Roebuck, J. The Negro numbers man as a criminal type: The construction and application of a typology. Journal of Criminal Law, Criminology and Police Science, 1963, 54, 48-60.
- Roebuck, J. H. Criminal typology. Springfield, Ill.: C. C. Thomas Publishers, 1967.
- Rulon, P. J. A simplified procedure for determining the reliability of a test by split halves. Harvard Educational Review, 1939, 9, 97-103.
- Samenow, D. The criminal personality. New York: Wiley Publishers, 1978.
- Schafer, S. The criminal and his victim. New York, N.Y.: Random House, 1968.
- Schafer, S. Theories in criminology. New York, N.Y.: Random House Publishers, 1969.
- Schlapp, M. G. The new criminology: A consideration of the chemical causes of abnormal behavior. New York: N.Y.: Boni and Liverright, 1928.
- Schrag, C. B. A preliminary criminal typology. Pacific Sociological Review, 1961, 4, 11-16.
- Schrag, C. The correctional system: Problems and Prospects. The Annals, 1969, 11-20.

- Schulman, W. J. Personality and behavior characteristics of assaultive patients (Doctoral dissertation, University of Minnesota, 1975). Dissertation Abstracts International, 1975, DAI 30:5695B.
- Schulsinger, F. Psychopathy, heredity and environment. International Journal of Mental Health, 1972, 22, 190-206.
- Sechrest, L. Incremental validity: A recommendation. Educational and Psychological Measurement, 1963, 23, 153-158.
- Sheldon, W. H. Atlas of man. New York, N.Y.: Harper and Row, 1949.
- Sheldon, W. H. Varieties of delinquent youth: An introduction to constitutional psychiatry. New York, N.Y.: Harper and Row, 1954.
- Shoham, S.; Gutmann, L.; & Rahav, G. A two-dimensional space for classification of legal offenses. Journal of Research in Crime and Delinquency, July 1970, 219-243.
- Sokal, R. R. Classification: Purposes, principles, progress, prospects. Science, 1974, 185(4157), 1115-1123.
- Stanley, J. C. Reliability. In R. L. Thorndike (Ed.), Educational Measurement (Rev. ed.). Washington, D.C.: American Council in Education, 1969.
- Stein, K. B.; Vadum, A. C.; & Serbin, T. Socialization and delinquency: A study of false positive and false negatives in prediction. Psychological Record, 1970, 20(3), 353-364.
- Stewart, D. K., & Love, W. A. A general canonical correlation index. Psychological Bulletin, 1968, 70, 160-163.
- Sutherland, E. H., & Cressey, D. R. Principals of criminology (7th ed.). Philadelphia, Penn.: Lippincott and Co., 1966.
- Tappan, P. W. Who is the criminal? American Sociological Review, 1967, 12, 96-102.
- Thrasher, F. M. The gang. Chicago, Ill.: University of Chicago Press, 1963.

- Timm, N. J. Multivariate analysis with applications in education and psychology. Bellmont, Calif.: Woods-worth Publishers, 1975.
- Vald, G. B. Theoretical criminology. New York, N.Y.: Oxford University Press, 1958.
- Warren, M. Q. Classification of offenders: An aid to efficient management and effective treatment. Journal of Research in Crime and Delinquency, 1969, 62, 239-291.
- Widom, C. S. An empirical classification of female offenders. Criminal Justice and Behavior, 1978, 8(1), 35-48.
- Wilkins, L. T., & Smith, P. M. Predictive attribute analysis. In N. Johnston, L. Savitz, & M. Wolfgang (Eds.), The sociology of punishment and correction (2nd ed.). New York: John Wiley and Sons, 1974.

BIOGRAPHICAL SKETCH

Brainard Willem Hines was born April 4, 1945, in Maxton, North Carolina. He moved with his parents to Charleston, West Virginia, in 1949, and attended elementary, junior high and high school in that city. He attended West Virginia University from 1963 until 1969, and obtained a Bachelor of Arts degree in psychology, as well as a Master of Science degree in clinical psychology.

From 1969 until 1974, he worked as a Program Evaluator at the Appalachia Educational Laboratory, and also was employed as Psychologist at the Charleston Guidance Clinic. He moved to Gainesville, Florida, to attend the University of Florida and to obtain a doctoral degree in foundations of education. He is currently residing in Miami, Florida.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William B. Ware
William B. Ware, Chairman
Professor of Foundations of
Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Linda Crocker
Linda Crocker
Professor of Foundations of
Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard M. Swanson
Richard M. Swanson
Professor of Psychology

This dissertation was submitted to the Graduate Faculty of the Department of Foundations of Education in the College of Education and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

June 1980

Robert P. Skerrett
Chairman, Foundations of Education

Francis J. Stebbins
Dean, Graduate School